

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPSrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on:      Mon Dec 27 11:37:08 1999;      MasPar time 2173.47 Seconds
Tabular output not generated.                1374.544 Million cell updates/sec

```

```

Title: >US-09-415-540-2
Description: (1-1275) from US09415540.seq
Perfect Score: 1274
N.A. Sequence:
Comp: 1 CACAGAGGTTNGSGGCTCTCT...CAATAAATAATTATTTGCTG 1275
      GTTCTCCANCCGCCGAGAGA.....CTTATTTTAAATAAAACAC

```

Scoring table: TABLE default

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: emb1-est58

Database: genbank-est111

8:9b-est1 9:9b-est10 10:9b-est11 11:9b-est12 12:9b-est13
13:9b-est14 14:9b-est15 15:9b-est16 16:9b-est17
17:9b-est18 18:9b-est19 19:9b-est20 20:9b-est21
21:9b-est22 22:9b-est23 23:9b-est24 24:9b-est25
25:9b-est26 26:9b-est27 27:9b-est28 28:9b-est29
29:9b-est30 30:9b-est31 31:9b-est32 32:9b-est33
33:9b-est34 34:9b-est35 35:9b-est36 36:9b-est37
37:9b-est38 38:9b-est39 39:9b-est40 40:9b-est41
41:9b-est42 42:9b-est43 43:9b-est44 44:9b-est45
45:9b-est46 46:9b-est47 47:9b-est48 48:9b-est49
49:9b-est50 50:9b-est51 51:9b-est52 52:9b-est53
53:9b-est54 54:9b-est55 55:9b-est56 56:9b-est57
57:9b-est58 58:9b-est59 59:9b-est60 60:9b-est61
61:9b-est62 62:9b-est63 63:9b-est64 64:9b-est65
65:9b-est66 66:9b-est67 67:9b-est68 68:9b-est69
69:9b-est70 70:9b-est71 71:9b-est72 72:9b-est73
73:9b-est74 74:9b-est75 75:9b-est76 76:9b-est77
77:9b-est78 78:9b-est79 79:9b-est80 80:9b-est81
81:9b-est82 82:9b-est83 83:9b-est84 84:9b-est85
85:9b-est86 86:9b-est87 87:9b-est88 88:9b-est89
89:9b-est90 90:9b-est91 91:9b-est92 92:9b-est93
93:9b-est94 94:9b-est95 95:9b-est96 96:9b-est97
97:9b-est98 98:9b-est99 99:9b-est100 100:9b-est101
101:9b-est102 102:9b-est103 103:9b-est104 104:9b-est105
105:9b-est106 106:9b-est107 107:9b-est108 108:9b-est109
109:9b-est110 110:9b-est111 111:9b-est112 112:9b-est113
113:9b-est114 114:9b-est115 115:9b-est116 116:9b-est117
117:9b-est118 118:9b-est119 119:9b-est120 120:9b-est121
121:9b-est122 122:9b-est123 123:9b-est124 124:9b-est125
125:9b-est126 126:9b-est127 127:9b-est128 128:9b-est129
129:9b-est130 130:9b-est131 131:9b-est132 132:9b-est133
133:9b-est134 134:9b-est135 135:9b-est136 136:9b-est137
137:9b-est138 138:9b-est139 139:9b-est140 140:9b-est141
141:9b-est142 142:9b-est143 143:9b-est144 144:9b-est145
145:9b-est146 146:9b-est147 147:9b-est148 148:9b-est149
149:9b-est150 150:9b-est151 151:9b-est152 152:9b-est153
153:9b-est154 154:9b-est155 155:9b-est156 156:9b-est157
157:9b-est158 158:9b-est159 159:9b-est160 160:9b-est161
161:9b-est162 162:9b-est163 163:9b-est164 164:9b-est165
165:9b-est166 166:9b-est167 167:9b-est168 168:9b-est169
169:9b-est170 170:9b-est171 171:9b-est172 172:9b-est173
173:9b-est174 174:9b-est175 175:9b-est176 176:9b-est177
177:9b-est178 178:9b-est179 179:9b-est180 180:9b-est181
181:9b-est182 182:9b-est183 183:9b-est184 184:9b-est185
185:9b-est186 186:9b-est187 187:9b-est188 188:9b-est189
189:9b-est190 190:9b-est191 191:9b-est192 192:9b-est193
193:9b-est194 194:9b-est195 195:9b-est196 196:9b-est197
197:9b-est198 198:9b-est199 199:9b-est200 200:9b-est201
201:9b-est202 202:9b-est203 203:9b-est204 204:9b-est205
205:9b-est206 206:9b-est207 207:9b-est208 208:9b-est209
209:9b-est210 210:9b-est211 211:9b-est212 212:9b-est213
213:9b-est214 214:9b-est215 215:9b-est216 216:9b-est217
217:9b-est218 218:9b-est219 219:9b-est220 220:9b-est221
221:9b-est222 222:9b-est223 223:9b-est224 224:9b-est225
225:9b-est226 226:9b-est227 227:9b-est228 228:9b-est229
229:9b-est230 230:9b-est231 231:9b-est232 232:9b-est233
233:9b-est234 234:9b-est235 235:9b-est236 236:9b-est237
237:9b-est238 238:9b-est239 239:9b-est240 240:9b-est241
241:9b-est242 242:9b-est243 243:9b-est244 244:9b-est245
245:9b-est246 246:9b-est247 247:9b-est248 248:9b-est249
249:9b-est250 250:9b-est251 251:9b-est252 252:9b-est253
253:9b-est254 254:9b-est255 255:9b-est256 256:9b-est257
257:9b-est258 258:9b-est259 259:9b-est260 260:9b-est261
261:9b-est262 262:9b-est263 263:9b-est264 264:9b-est265
265:9b-est266 266:9b-est267 267:9b-est268 268:9b-est269
269:9b-est270 270:9b-est271 271:9b-est272 272:9b-est273
273:9b-est274 274:9b-est275 275:9b-est276 276:9b-est277
277:9b-est278 278:9b-est279 279:9b-est280 280:9b-est281
281:9b-est282 282:9b-est283 283:9b-est284 284:9b-est285
285:9b-est286 286:9b-est287 287:9b-est288 288:9b-est289
289:9b-est290 290:9b-est291 291:9b-est292 292:9b-est293
293:9b-est294 294:9b-est295 295:9b-est296 296:9b-est297
297:9b-est298 298:9b-est299 299:9b-est300 300:9b-est301
301:9b-est302 302:9b-est303 303:9b-est304 304:9b-est305
305:9b-est306 306:9b-est307 307:9b-est308 308:9b-est309
309:9b-est310 310:9b-est311 311:9b-est312 312:9b-est313
313:9b-est314 314:9b-est315 315:9b-est316 316:9b-est317
317:9b-est318 318:9b-est319 319:9b-est320 320:9b-est321
321:9b-est322 322:9b-est323 323:9b-est324 324:9b-est325
325:9b-est326 326:9b-est327 327:9b-est328 328:9b-est329
329:9b-est330 330:9b-est331 331:9b-est332 332:9b-est333
333:9b-est334 334:9b-est335 335:9b-est336 336:9b-est337
337:9b-est338 338:9b-est339 339:9b-est340 340:9b-est341
341:9b-est342 342:9b-est343 343:9b-est344 344:9b-est345
345:9b-est346 346:9b-est347 347:9b-est348 348:9b-est349
349:9b-est350 350:9b-est351 351:9b-est352 352:9b-est353
353:9b-est354 354:9b-est355 355:9b-est356 356:9b-est357
357:9

Statistics: Mean 11.461; Variance 2.065; scale, 5.550

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
c	1	710	55.7	751	22	AA078153		oz30c03.x1 Soares_feta	0.00e+00
c	2	666	49.1	639	17	AA694204		z42c12.x1 Soares_feta	0.00e+00
c	3	596	46.8	704	28	A1656500		tol6ef09.x1 NCI_CGAP_Ut	0.00e+00
c	4	571	44.8	646	18	AA766443		oa35e06.s1 NCI_CGAP_GC	0.00e+00
c	5	567	44.5	570	24	A1215898		qk35f09.x1 NCI_CGAP_Lu	0.00e+00
c	6	541	42.5	564	14	AA493445		ne48f03.s1 NCI_CGAP_Co	0.00e+00
c	7	529	41.5	540	20	AA932061		co059b07.s1 NCI_CGAP_Lu	0.00e+00
c	8	529	41.5	638	36	AA075389		zm87c06.s1 StrataGene	0.00e+00
c	9	523	41.1	572	24	A1219448		qhl3c03.x1 Soares_NFL	0.00e+00
c	10	520	40.8	530	18	AA777776		z195f10.s1 Soares_feta	0.00e+00

C	11	519	40.7	521.22	AA1096370	qb91e02.x1	Soares_feta	0.00e+00
C	12	518	40.7	528.11	AA314881	ac34e04.s1	Stratagene	0.00e+00
C	13	518	40.7	572.16	AA634465	ac34e04.s1	Stratagene	0.00e+00
C	14	513	40.3	521.17	AA702133	z185908.x1	Soares_feta	0.00e+00
C	15	514	40.3	580.36	AA075548	z185908.x1	Stratagene	0.00e+00
C	16	507	39.8	512.21	AA935739	o079h02.s1	CGAP_Pt	0.00e+00
C	17	505	39.8	538.16	AA585751	nr83k04.s1	NCI_CGAP_R1	0.00e+00
C	18	500	39.2	531.20	AA864283	ch54e04.s1	NCI_CGAP_GC	0.00e+00
C	19	484	38.0	506.18	AA779373	z157e07.s1	Soares_feta	0.00e+00
C	20	482	37.8	493.18	AA836526	ap21f06.s1	NCI_CGAP_GC	0.00e+00
C	21	471	37.0	475.24	AI1772200	ap21f10.x1	NCI_L1iber	0.00e+00
C	22	467	36.7	471.20	AA878922	o125d10.s1	NCI_CGAP_R1	0.00e+00
C	23	467	36.7	475.14	AA502824	ne42h11.s1	NCI_CGAP_Co	0.00e+00
C	24	468	36.7	480.20	AA483368	aj16a10.s1	Soares_Para	0.00e+00
C	25	464	36.4	466.22	AI1088393	qb11c06.x1	Soares_Preg	0.00e+00
C	26	462	36.3	479.11	AA316013	EST188096	HCC cell 11n	0.00e+00
C	27	451	35.4	467.36	AI102865	tn80f11.s1	Stratagene	0.00e+00
C	28	450	35.3	492.28	AI521391	ti05h05.x1	NCI_CGAP_Pa	0.00e+00
C	29	449	35.2	457.10	AA779355	zS84e07.r1	NCI_CGAP_GC	0.00e+00
C	30	446	35.2	483.15	AA588711	nm17h03.s1	NCI_CGAP_Co	0.00e+00
C	31	448	35.0	446.11	AA314838	EST186421	Colo. carc. in	0.00e+00
C	32	444	34.9	465.17	AA7444991	nv43b02.s1	NCI_CGAP_Ev	0.00e+00
C	33	441	34.6	458.17	AA730829	nv49f03.s1	NCI_CGAP_Ev	0.00e+00
C	34	437	34.3	449.21	AA430893	op22b11.s1	NCI_CGAP_Co	0.00e+00
C	35	436	34.2	437.34	W67406.37	zdf0f04.r1	Soares_feta	0.00e+00
C	36	436	34.2	470.15	AA608572	ae54e07.s1	Stratagene	0.00e+00
C	37	435	34.1	653.35	AA030742	m131f01.r1	Soares_mous	0.00e+00
C	38	429	33.7	482.24	AI269018	q173b09.x1	NCI_CGAP_R1	0.00e+00
C	39	428	33.6	428.36	AA102844	zn80f11.r1	Stratagene	0.00e+00
C	40	425	33.4	461.25	AI283314	qui4a03.x1	NCI_CGAP_Ov	0.00e+00
C	41	420	33.0	520.31	H50229	yo28c01.s1	Soares_adul	0.00e+00
C	42	419	32.9	481.25	AI351161	qt11h09.x1	NCI_CGAP_GC	0.00e+00
C	43	415	32.6	591.9	AA213035	nm79f06.r1	Soares_mous	0.00e+00
C	44	411	32.3	446.36	AA075325	zm86a12.s1	Stratagene	0.00e+00
C	45	412	32.3	569.14	AA558853	vk10e12.r1	Knowles_Sol	0.00e+00

ALIGNMENTS

RESULT	1	AI078153	751 bp	mRNA	EST	01-OCT-1998
LOCUS		oz20c03.x1	Seares-total_fetus.NB2HF6.9w	Homo sapiens	CDNA clone	
DEFINITION		IMAGE:1676836.3	similar to SW:IPR_BOVIN_P37360	INORGANIC PYROPHOSPHATASE	;', mRNA sequence.	
ACCESSION		AI078153				
NID		93412561				
VERSION		AI078153.1	GI:3412561			
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
		Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
REFERENCE		1. (bases 1 to 751)				
AUTHORS		NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CCAP).				
		Tumor Gene Index				
JOURNAL		Unpublished (1997)				
COMMENT		On Jan 19, 1998 this sequence version replaced gi:2282347.				

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNCX; contact the
IMAGE Consortium (info@image.lni.gov) for further information.
Insert Length: 2041 Std Error: 0.00
Seq primer: -40m13 fwd: EP from Amersham
High quality sequence stop: 452.

/organism="Homo sapiens"
/note="Vector: pT7SD-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(5') primer [5']
 TGTTCACATCTGAGCTGGAGCGCGCTTATATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTZ19 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

/db_xref="taxon:9606"
 /clone="IMAGE:1676836"
 /clone_lib="Soares_fetal_liver_spleen_INFIS_S1"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"

BASE COUNT 222 a 151 c 121 g 257 t

Query Match 55.7%; Score 710; DB 22; Length 751;
 Best Local Similarity 98.5%; Pred. No. 0.00e+00;
 Matches 731; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

Db 10 CAGCAAAATATTTTTCCTACATATGATATACATATACATATGCTCCCT 69
 |||||||
 Cp 1275 CAGCAAAATATTTTTCCTACATATGATATACATATACATATGCTCCCT 1216
 |||||||
 Db 70 TGCACATCTATTCACAGTACCTCCAAATGACAACTGCTTGATATTAACATGTGT 129
 |||||||
 Cp 1215 TGCACATCTATTCACAGTACCTCCAAATGACAACTGCTTGATATTAACATGTGT 1156
 |||||||
 Db 130 AAAAGTATCTAGTTGAGATATGAAAAATGCTTTAATGATATACATCTGATATAT 189
 |||||||
 Cp 1155 AAAAGTATCTAGTTGAGATATGAAAAATGCTTTAATGATATACATCTGATATAT 1096
 |||||||
 Db 190 GGATAGTACAGCAAAATTTCTAGTATGATGCTCTCAAAATTAAGCTTTGAA 249
 |||||||
 Cp 1095 GGATAGTACAGCAAAATTTCTAGTATGATGCTCTCAAAATTAAGCTTTGAA 1036
 |||||||
 Db 250 AAGCTACTACTTTTACTCTTAATACATCCAGATGACAGATGATACATACCTGT 309
 |||||||
 Cp 1035 AAGCTACTACTTTTACTCTTAATACATCCAGATGACAGATGATACATACCTGT 976
 |||||||
 Db 310 ATTCCAGAGAAATCTCATTTAGTTTTCGTGATATGAAACCTTTTCCAGCTGTGTGT 369
 |||||||
 Cp 975 ATTCCAGAGAAATCTCATTTAGTTTTCGTGATATGAAACCTTTTCCAGCTGTGTGT 916
 |||||||
 Db 370 ACTGTCAGGAGATTCACAGGCTGTGTAAGATCCACAAATGGCTGTGGACATCA 429
 |||||||
 Cp 915 ACTGTCAGGAGATTCACAGGCTGTGTAAGATCCACAAATGGCTGTGGACATCA 856
 |||||||
 Db 430 GGATACACTGGAAGGGCTCTCAAAACAAAGTTTATTCATGCAACTGATTCCTTTCCA 489
 |||||||
 Cp 855 GGATACACTGGAAGGGCTCTCAAAACAAAGTTTATTCATGCAACTGATTCCTTTCCA 796
 |||||||
 Db 490 TTGCTTTCTTAGTACATTAATGCTTTCAATGTCATAGAGCTTTTAATATATCATG 549
 |||||||
 Cp 795 TTGCTTTCTTAGTACATTAATGCTTTCAATGTCATAGAGCTTTTAATATATCATG 736
 |||||||
 Db 550 GCAAGTCTTATCTTTAATTCGATTAACGCAAACTATTTTCGTTTCCATCA 609
 |||||||
 Cp 735 GCAAGTCTTATCTTTAATTCGATTAACGCAAACTATTTTCGTTTCCATCA 676
 |||||||
 Db 610 GGAACCTATACCTTTTAACGAGTCCACATAGCTTTCTAGTAGCCAGGTTTCAACC 669
 |||||||
 Cp 675 GGAACCTATACCTTTTAACGAGTCCACATAGCTTTCTAGTAGCCAGGTTTCAACC 618
 |||||||
 Db 670 GTTTCACATCTATGATATCATATTAATGCTGATGATGATATTCACATTAAGCA 729
 |||||||
 Cp 617 GTTTCACATCTATGATATCATATTAATGCTGATGATGATATTCACATTAAGCA 558
 |||||||
 Db 730 TGACTTTCCATGCGTTTCCC 751;
 |||||||
 Cp 557 TGACTTTCCATGCGTTTCCC 536;
 |||||||

RESULT 2

LOCUS AA694204 639 bp mRNA EST 16-DEC-1997
 DEFINITION z142c12.s1 Soares fetal liver spleen INFIS_S1 Homo sapiens CDNA
 clone IMAGE:433462 3' similar to SW-IPR_BOVIN p37960 INORGANIC
 PYROPHOSPHATASE ;, mRNA sequence.
 ACCESSION AA694204
 NID 92695142
 VERSION AA694204.1 GI:2695142
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 639)
 Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kueba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
 Martin,J., Moore,B., Schellenberg,R., Steptoe,M., Tan,F.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)
 On Nov 12, 1997 this sequence version replaced gi:2559131.
 COMMENT
 TITLE JOURNAL
 JOURNAL

FEATURES
 source
 1. 639
 /organism="Homo sapiens"
 /note="Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia)
 with a modified polylinker. Site.1: Pac I; Site.2: Eco RI;
 This is a subtracted version of the original Soares fetal
 liver spleen INFIS library. 1st strand cDNA was primed
 with a Pac I - oligo(5') primer [5']
 AACTGAGAAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pTZ19 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 /db_xref="GDB:1333231"
 /db_xref="taxon:9606"
 /map="12"
 /clone="IMAGE:433462"
 /clone_lib="Soares_fetal_liver_spleen_INFIS_S1"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"

BASE COUNT 192 a 126 c 107 g 214 t

Query Match 49.1%; Score 626; DB 17; Length 639;
 Best Local Similarity 99.1%; Pred. No. 0.00e+00;
 Matches 632; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 CAGCAAAATATTTTTCCTACATATGATATACATATACATATGCTCCCT 61
 |||||||
 Cp 1275 CAGCAAAATATTTTTCCTACATATGATATACATATACATATGCTCCCT 1216
 |||||||
 Db 62 TGCACATCTATTCACAGTACCTCCAAATGACAACTGCTTGATATTAACATGTGT 121
 |||||||
 Cp 1215 TGCACATCTATTCACAGTACCTCCAAATGACAACTGCTTGATATTAACATGTGT 1156
 |||||||
 Db 122 AAAAGTATCTAGTTGAGATATGAAAAATGCTTTAATGATATACATCTGATATAT 181
 |||||||
 Cp 1155 AAAAGTATCTAGTTGAGATATGAAAAATGCTTTAATGATATACATCTGATATAT 1096
 |||||||

```

/organism="Homo sapiens"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.85 kb. Life Technologies catalog #

```

SOURCE

TTTACTTCTAATACATCCAGATGAAACCATCTTCCCAATATCTACCCCTCT 076

FEATURES	SOURCE
<p>Db 342 ATTCAGAGAAATCCATAGTATTTTCGGATGAGCAACACTATCCAGCGCTTGGT 401</p> <p>Cp 975 ATTCAGAGAAATCCATAGTATTTTCGGATGAGCAACACTATCCAGCGCTTGGT 916</p> <p>Db 402 ACTGTGCAGGCAGATTCACAGGGGTGGTGGTAAAGCATCCACATGGCTCTGGAGCATCA 461</p> <p>Cp 915 ACTGTGCAGGCAGATTCACAGGGGTGGTGGTAAAGCATCCACATGGCTCTGGAGCATCA 856</p> <p>Db 462 GGATCACACTTGAGGGGGGCTCTCAGACAAAGGTGGTATCCATGACATGATCTTCTGCCA 521</p> <p>Cp 855 GGATCACACTTGAGGGGGGCTCTCAGACAAAGGTGGTATTCATGACATGATCTTCTGCCA 796</p> <p>Db 532 TTCGTTTCTAGGGCCCTTAAGGCTTTTCCAAATGGTCATGATGCTTTTAATTAATATACATG 581</p> <p>Cp 795 TTCGTTTCTAGGGCCCTTAAGGCTTTTCCAAATGGTCATGATGCTTTTAATTAATATACATG 736</p> <p>Db 582 GCAAGTCTTATCTTAATCTTAATTTGTCATTAACCCGCAAACTCATTTTCGGTTTTCATCA 641</p> <p>Cp 735 GCAAGTCTTATCTTAATCTTAATTTGTCATTAACCCGCAAACTCATTTTCGGTTTTCATCA 676</p> <p>Db 642 GGAC 646</p> <p>Cp 675 GGAC 671</p>	<p>RESULT 5</p> <p>LOCUS A1215898 570 bp mRNA EST 21-DEC-1998</p> <p>DEFINITION QM35H09.x1 NCI-CGAP L45 Homo sapiens CDNA clone IMAGE:1883873 3' similar to SW:EFYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA sequence.</p> <p>ACCESSION A1215898</p> <p>NID 93784939</p> <p>VERSION A1215898.1 GI:3784939</p> <p>KEYWORDS EST.</p> <p>SOURCE human.</p> <p>ORGANISM Homo sapiens</p> <p>REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.</p> <p>AUTHORS I (bases 1 to 570)</p> <p>TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.</p> <p>COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index</p> <p>JOURNAL Unpublished (1997)</p> <p>COMMENT On Jan 14, 1998 this sequence version replaced gi:1796996.</p>

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bdip/image/image.html

Insert Length: 1001 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 353.
Location/Qualifiers
1. 570
/organism="Homo sapiens"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"
/clone="IMAGE:1883873"

Cp 735 G 735

RESULT 7 AA932061 540 bp mRNA EST 07-JUL-1998
LOCUS 0059b07.s1 NCI_CGAP_Lus Homo sapiens cDNA clone IMAGE:1570453 3'
DEFINITION similar to SW:IPR_BOVIN_P37980 INORGANIC PYROHOSPHATASE ;, mRNA
sequence
ACCESSION AA932061
NID 93086374
VERSION AA932061.1 GI:3086374
KEYWORDS EST
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 540)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402299.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnlnl.gov/bdrip/image/image.html

Insert Length: 1022 Std Error: 0.00
Seq primer: -40ml3 fwd. fr. from Amersham
High quality sequence stop: 395.

FEATURES

Location/Qualifiers

1..340

/organism="Homo sapiens"

/note="Organ: lung; Vector: pRT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from a
neuroendocrine lung carcinoma, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
pRT3 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"

/clone_id="NCI_CGAP_Lus"

/tissue_type="carcinoid"

/lab_host="DH10B"

BASE COUNT 164 a 103 c 94 g 179 t

ORIGIN

Query Match 41.5%; Score 529; DB 20; Length 540;
Best Local Similarity 99.1%; Pred. 0.00e+00;
Matches 534; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 2 CAGCAAAATATTTATTTCTTACATATGATATATTCATATGATGCTCCCT 61
Cp 1275 CAGCAAAATATTTATTTCTTACATATGATATATTCATATGATGCTCCCT 1216
Db 62 TGCACATATTCACAGTACCTTCAATGACAACTGCTTTGATTTAGCATGTCT 121
Cp 1215 TGCACATATTCACAGTACCTTCAATGACAACTGCTTTGATTTAGCATGTCT 1156
Db 122 AAAAGTATCTTGTAGATATGAAAATGCTTTATATGATATATTTAGATATTT 181
Cp 1155 AAAAGTATCTTGTAGATATGAAAATGCTTTATATGATATATTTAGATATTT 1096

Db 182 GGATTAGTACAGCAGAAATTTACTTATGATGAGTCTTACAAATTTAAGCTTGA 241
Cp 1095 GGATTAGTACAGCAGAAATTTACTTATGATGAGTCTTACAAATTTAAGCTTGA 1036
Db 242 AAGCTACTACTTTTACTTCTTATATATCATTCCAGATGACAGATGAGCAATATACGTTGT 301
Cp 1035 AAGCTACTACTTTTACTTCTTATATATCATTCCAGATGACAGATGAGCAATATACGTTGT 976
Db 302 ATTCCAGAAATCTCAATTTATTTTCTGGTATGAGAACCACTATCCAGCTGTGT 361
Cp 975 ATTCCAGAAATCTCAATTTATTTTCTGGTATGAGAACCACTATCCAGCTGTGT 916
Db 362 ACTGTGACGAGCATTCACAGGGTGTGTGTAAGCATCCAAATGGCTGGCAGCATCA 421
Cp 915 ACTGTGACGAGCATTCACAGGGTGTGTGTAAGCATCCAAATGGCTGGCAGCATCA 856
Db 422 GGATCACACTTGAAGGGCTCTCAGACAAAGTTGATTCATGCAACTGATTCCTTTCC 481
Cp 855 GGATCACACTTGAAGGGCTCTCAGACAAAGTTGATTCATGCAACTGATTCCTTTCCA 796
Db 482 TTGCTTTCCAGTCACTAATGCTTTCCATGTCATGAGTCTTTAATATATCAAT 540
Cp 795 TTGCTTTCTTACTACTAATGCTTTCCATGTCATGAGTCTTTAATATATCAAT 737

RESULT 8 AA075389 638 bp mRNA EST 23-DEC-1997
LOCUS zmb7c06.s1 StrataGene ovarian cancer (#937219) Homo sapiens cDNA
DEFINITION clone IMAGE:544906 3' similar to SW:IPR_BOVIN_P37980 INORGANIC
PYROHOSPHATASE ;, mRNA sequence.
ACCESSION AA075389
NID 91615260
VERSION AA075389.1 GI:1615260
KEYWORDS EST
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
WashU-Merck EST Project
Unpublished (1995)

JOURNAL

COMMENT On Apr 14, 1993 this sequence version replaced gi:716386.

Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewartson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnlnl.gov) for further information.
Seq primer: -40ml3 fwd. from Amersham

High quality sequence stop: 236.

FEATURES

source

1..638

/organism="Homo sapiens"

/note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Papillary serous carcinoma, isolated from ascites, 64 year
old caucasian. Average insert size: 0.8 kb; Uni-ZAP XR
Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3'
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

/db_xref="GDB:3923094"

Query Match	41.5%	Score 529	DB 36	Length 638
Best Local Similarity	96.6%	Pred. No. 0.00e+00	Mismatches 14	Indels 6
Matches 576	Conservative	0	Mismatches 14	Indels 6
Db	2	CAGCAAAATAATTTTATTTCCCTCAATATGCTTAACATATACATCCAAATATGTCCTCCCT	61	
Cp	1275	CAGCAAAATAATTTTATTTCCCTCAATATGCTTAACATATACATCCAAATATGTCCTCCCT	1216	
Db	62	TGCACATCTATTCACAGAGTGAAGTCTTCCAAATATACACATGCTTTGATATTTAAGCATGTGCT	121	
Cp	1215	TGCACATCTATTCACAGAGTGAAGTCTTCCAAATATACACATGCTTTGATATTTAAGCATGTGCT	1156	
Db	122	AAAAGTTATCTAGTGTGATGATGAAATATGCTTATGATGATATTAACATCTGAGATATAT	181	
Cp	1155	AAAAGTTATCTAGTGTGATGATGAAATATGCTTATGATGATATTAACATCTGAGATATAT	1096	
Db	182	GGATTAGTCACAGCAGAAATTTACTTACTAGATGATGATTTCAAAATTTAAAGCTTTGA	241	
Cp	1095	GGATTAGTCACAGCAGAAATTTACTTACTAGATGATGATTTCAAAATTTAAAGCTTTGA	1037	
Db	242	AAAGCTACTACTTTTACTCTTAATATCATCCAGATGACAGATGTAGCATATATACGTTG	301	
Cp	1036	AAAGCTACTACTTTTACTCTTAATATCATCCAGATGACAGATGTAGCATATATACGTTG	977	
Db	302	TATTCACAGAAATCTCATGTTTCTTGATGATGAGAACACATTCACAGCTGTGTTG	361	
Cp	976	TATTCACAGAAATCTCATGTTTCTTGATGATGAGAACACATTCACAGCTGTGTTG	917	
Db	362	TACTGTGAGGCACAGATTCACAGAGGTGTGTAAAGCATCCAAATGGGCTGTGCACAT	421	
Cp	916	TACTGTGAGGCACAGATTCACAGAGGTGTGTAAAGCATCCAAATGGGCTGTGCACAT	858	
Db	422	CAGGATCACACTTGAAGGGGCTCTCAGACAAGTTGGTATTCATGACAATGATTCCTTT	481	
Cp	857	CAGGATCACACTTGAAGGGGCTCTCAGACAAGTTGGTATTCATGACAATGATTCCTTT	799	
Db	482	CCAAATCGTTTCAGTACTAATAGCTTCCAAATGGGATAGAGCCCTTAATATATCA	541	
Cp	798	CCAAATCGTTTCAGTACTAATAGCTTCCAAATGGGATAGAGCCCTTAATATATCA	739	
Db	542	ATGGGCAAAAGTCCCTATCCTTAATATCNGCATTTAA-CGCA-CCTCATTTGCGGTT	595	
Cp	738	ATGG-CAAAGTCCCTATCCTTAATATCNGCATTTAAAGCACAATCATTTTCTGGTT	684	
RESULT	9			
LOCUS	A1219448	572 bp	mRNA	EST
DEFINITION	g113c03.x1 Soares.NFL.T.GBC.S1 Homo sapiens cDNA clone			30-NOV-1998
	IMAGE:1844548.3	similar to SM:IPYR_BOVIN P37980	INORGANIC	
	PYROPHOSPHATASE ;	mRNA sequence.		
ACCESSION	A1219448			
NID	g3801651			
VERSION	A1219448.1	GI:3801651		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .			
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			

On Aug 21, 1998 this sequence replaced

```

/organism="Homo sapiens"
/notice="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were used as
in vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The drive
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries (The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
/db_xref="taxon:9606"
/clone="IMAGE:1844548"
/clone_lib="Soares_NFL_T-GBC-S1"
/lab_host="DH10B"
1173 a 109 c 97 g 192 t 1 others
BASE COUNT
ORIGIN

```

	Query Match:	41.1%	Score 523:	DB 24:	Length 572:
	Best Local Similarity	97.6%:	Pred. No.	0.00e+00:	
	Matches 561:	Conservative	0:	Mismatches 9:	Indels 5: Gaps 5:
Db	3	CAGCAAAATAATTTTATTTCCTTAACATATGTAACAATATCATCCATATATGTGTCCTCCT	62		
Cp	1275	CAGCAAAATAATTTTATTTCCTTAACATATGTAACAATATCATCCATATATGTGTCCTCCT	1216		
Db	63	TGCACATCTATTTCACAAAGTGACTGCCAATGACAACTGCTTTGATATTTAAGCATGTGCT	122		
Cp	1215	TGCACATCTATTTCACAAAGTGACTGCCAATGACAACTGCTTTGATATTTAAGCATGTGCT	1156		
Db	123	AAAAGTTATCTTAAGTAGATGAATGAAAAAATGCTTTAAATGGATTAACCTTGCGATATT	182		
Cp	1155	AAAAGTTATCTTAAGTAGATGAATGAAAAAATGCTTTAAATGGATTAACCTTGCGATATT	1096		
Db	183	GGATTAGTCACAGCAGAATTTACTTTAGTAGAGAGTTCTACAAATTTAAAGCTTGA	242		
Cp	1095	GGATTAGTCACAGCAGAATTTACTTTAGTAGAGAGTTCTACAAATTTAAAGCTTGA	1036		
Db	243	AAGCTACTACTTTTACTTTCTTAATTCATCCACAGATGACACAGATGTAGCAATACGTTGT	302		
Cp	1035	AAGCTACTACTTTTACTTTCTTAATTCATCCACAGATGACACAGATGTAGCAATACGTTGT	976		
Db	303	ATTCAGAGAAATCTCAATTTAGTTTCTGGATGATGSAACCACTATCCAGCTCGTTGGT	352		
Cp	975	ATTCAGAGAAATCTCAATTTAGTTTCTGGATGATGSAACCACTATCCAGCTCGTTGGT	916		
Db	363	ACTGTGCGAGCAGATTTCACAGAGGTGGTGG- AAAGCATCCCAATGGCTGTGGCAGCATCA	421		
p	915	ACTGTGCGAGCAGATTTCACAGAGGTGGTGGTAAAGCATCCCAATGGCTGTGGCAGCATCA	856		
Db	422	GGATTCACACTTTGAAGGGGCTCTCAGACAAAGTTGTATTCAATGCAACTGATTCCTTTTCCA	481		
Cp	855	GGATTCACACTTTGAAGGGGCTCTCAGACAAAGTTGTATTCAATGCAACTGATTCCTTTTCCA	796		
Db	482	T-CGTTTTCTTA-TCACTAATGCTTTCACAGTGNACATGA-TGCTTTTA-TAATATCAATG	537		
Cp	795	TTGCTTTTCTTAAGTCACTTAATGCTTTCACAGTGNACATGAAGTGAAGTCTTTTAATATATCAATG	736		
Db	538	GCAAAGTCCTTTATCTTTAATTTGCAATAAAGCG	572		

	Query Match	40.8%	Score 520	DB 18	Length 530
	Best Local Similarity	99.1%	Pred. No. 0.00e+00		
	Matches 525	Conservative	0	Mismatches 5	Indels 0
				Gaps	0
db	1	CAGCAAAATATTTTATTTTCTTAACATATGTGTAACATATATCAATCATATGTGCTCCCT	60		
db	1275	CAGCAAAATATTTTATTTTCTTAACATATGTGTAACATATATCAATCATATGTGCTCCCT	1216		
db	61	TGCACATCTATTCACAAGTGCATCCCAATGAAACAGCGCTTATATTTAAAGCATGTGT	120		
db	1215	TGCACATCTATTCACAAGTGCATCCCAATGAAACAGCGCTTATATTTAAAGCATGTGT	1156		

Db	1321	AAAGTTATCTTACTAGATATCAAAAAGCTTTGATGAGTAACATCTCGATATAT	180
Cp	1135	AAAATTATCTTACTAGATATGAAAAAGCTTTTGATGAGTAACATCTCGATATAT	1096
Db	181	GGATTAGTCACAGCAGATTACTTTAGTAGATGAGTCTCAAAATTTAAAGCTTGA	240
Cp	1095	GGATTAGTCACAGCAGAAATTTACTTAGTAGATGAGTCTCAAAATTTAAAGCTTGA	1033
Db	241	AAGCTACTACTTTACTTTCTAATATCATCCAGATGAACAGATGTAGCAATATACGTTGT	300
Cp	1035	AAGCTACTACTTTACTTTCTAATATCATCCAGATGAACAGATGTAGCAATATACGTTGT	976
Db	301	ATTCACAGAGAAATCTCTATTAGTTTTCGTGGTATGGAACACCTPATTCACGCTTGST	360
Cp	975	ATTCACAGAGAAATCTCTATTAGTTTTCGTGGTATGGAACACCTPATTCACGCTTGST	916
Db	361	ACTGTGACGAGCATTTCCACAGSGTGTGTTAAAGCATCCAAATGGCTGTGCACATCA	420
Cp	915	ACTGTGACGAGCATTTCCACAGSGTGTGTTAAAGCATCCAAATGGCTGTGCACATCA	856
Db	421	GGATTCACACTTGAAGGGGGCTTCAGACAAAGTTGTATTATGCAACTGATTCCTTTCCA	480
Cp	855	GGATTCACACTTGAAGGGGGCTTCAGACAAAGTTGTATTATGCAACTGATTCCTTTCCA	796
Db	481	ATCGTTTCTTACTACTATATGCGCTTCCAAATGATATGATGTCCTTTAT	530
Cp	795	ATCGTTTCTTACTACTATATGCGTTTCCAAATGATATGATGTCCTTTAT	746

```

RESULT 11
LOCUS 521 bp mRNA EST
DEFINITION gp91e02.x1 Soares_fetal_heart_NbHH19w Homo sapiens CDNA clone
IMAGE:1070482.3 similar to SW:IPYR_BOVIN P37980 INORGANIC
PYROPHOSPHATASE / mRNA sequence.
ACCESSION AI096370
NID G3446281
VERSION AI096370.1 GI:3446281
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 521)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
COMMENT Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1400967.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: 40m13 fwd, ET from Amersham
High quality sequence stop: 452.
Location/Qualifiers
1..521
FEATURES
source

```

```

/organism="Homo sapiens"
/notice="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(4T) primer
TGTTCACCAATCTGATAGTGAGCGCGCCCATCTTTTTTTTTTTTTTTT
3')
double-stranded cDNA was size selected, ligated to Eco
adapters (Pharmacia), digested with Not I and cloned in
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fátima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL197."
/db_xref="taxon:9606"
/clone="IMAGE:1707482"

```


/clone_lib="Soares_fetal_heart_NbH19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 BASE COUNT 156 a 99 c 93 g 173 t
 ORIGIN

Query Match 40.7%; Score 519; DB 22; Length 521;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 520; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CAGCAAAATATTTATTTCTTACATATGTACATATACATATCAATATGTCCTCCCT 60
 Cp 1275 CAGCAAAATATTTATTTCTTACATATGTACATATACATATCAATATGTCCTCCCT 1216
 Db 61 TGCACATCTATTCACAGAGCTTCCAAATGACAACTGCTTGATATTTAAGCATGTCT 120
 Cp 1215 TGCACATCTATTCACAGAGCTTCCAAATGACAACTGCTTGATATTTAAGCATGTCT 1156
 Db 121 AAAATTTATCTAGTGTGATGATGAAATGCTTTAGATGATGATGATGATGATATAT 180
 Cp 1155 AAAATTTATCTAGTGTGATGATGAAATGCTTTAGATGATGATGATGATGATATAT 1096
 Db 181 GGATTTAGTACAGAGCAATTTACTTTAGTATGATGATGATGATGATGATGATGAT 240
 Cp 1095 GGATTTAGTACAGAGCAATTTACTTTAGTATGATGATGATGATGATGATGATGAT 1036
 Db 241 AAGTACTACTTTTACTTCTTATACATCCAGATGAGATGATGATGATGATGATGATGAT 300
 Cp 1035 AAGTACTACTTTTACTTCTTATACATCCAGATGAGATGATGATGATGATGATGATGAT 976
 Db 301 ATTCCAGGAATCTCATTTAGTGTGATGATGATGATGATGATGATGATGATGATGAT 360
 Cp 975 ATTCCAGGAATCTCATTTAGTGTGATGATGATGATGATGATGATGATGATGATGAT 916
 Db 361 ACTGTGACAGGAGATTTCACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 420
 Cp 915 ACTGTGACAGGAGATTTCACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 856
 Db 421 GGATACACTTGAAGGGCTCTCAAGCAAAAGTGTATTCATCACTGATGATGATGATGAT 480
 Cp 855 GGATACACTTGAAGGGCTCTCAAGCAAAAGTGTATTCATCACTGATGATGATGATGAT 796
 Db 481 TTCGTTTCTTTCATCAATGCTTTCCATGATGATGATGATGATGATGATGATGATGAT 521
 Cp 795 TTCGTTTCTTTCATCAATGCTTTCCATGATGATGATGATGATGATGATGATGATGAT 755

RESULT 12 AA314881 528 bp mRNA EST 19-APR-1997
 LOCUS EST186922 HCC cell line (metastasis to liver in mouse) II Homo
 DEFINITION sapiens cDNA 5' end similar to similar to inorganic
 pyrophosphatase, mRNA sequence.
 ACCESSION AA314881
 MID 91967451
 VERSION AA314881.1 GI:1967451
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 528)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulmer,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-
 Mait,C., Clontz,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
 Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,
 Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 On Apr 14, 1993 this sequence version replaced gi:785883.
 Other ESTs: THC195084
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers

1..528
 /organism="Homo sapiens"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"
 /db_xref="ATCC (Inhost):113250"
 /db_xref="taxon:9606"
 /map="844D06; 5; 5q31.1-5q31.3"
 /clone_lib="HCC cell line (metastasis to liver in mouse)
 II"
 /tissue-type="colon"
 /cell_type="KM12SM"
 /cell_line="KM12C(HCC)metastasis into mouse (liver)"
 BASE COUNT 155 a 101 c 140 g 132 t
 ORIGIN

Query Match 40.7%; Score 518; DB 11; Length 528;
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;
 Matches 526; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 TGGTGGCTCTGTGCGAGGGCGGGCGAGAGCTCCGCGCATATGAGCGGCTTCACACCG 60
 Cp 48 TGGTGGCTCTGTGCGAGGGCGGGCGAGAGCTCCGCGCATATGAGCGGCTTCACACCG 107
 Db 61 AGGAGCGGCGCGCGCC-TCCTCCCTGAGATACGAGTCTCTCTCAAAATGGAAGAGC 119
 Cp 108 AGGAGCGGCGCGCGCCCTTCTCCCTGAGATACGAGTCTCTCTCAAAATGGAAGAGC 167
 Db 120 AATATATCTTCCATTTTCATGATATTCGAATTTATGAGATGATGATGATGATGATGAT 179
 Cp 168 AATATATCTTCCATTTTCATGATATTCGAATTTATGAGATGATGATGATGATGATGAT 227
 Db 180 TAGTTGAAGTACACGCTGCTGCTAATGCAAAATGAGATGCTCAAGAGACCTTTAA 239
 Cp 228 TAGTTGAAGTACACGCTGCTGCTAATGCAAAATGAGATGCTCAAGAGACCTTTAA 287
 Db 240 ACCCTATTAACAAGATGTAAGAAAAAGAAAACTTCGATATTCGAAATTTGTTCCCT 299
 Cp 288 ACCCTATTAACAAGATGTAAGAAAAAGAAAACTTCGATATTCGAAATTTGTTCCCT 347
 Db 300 ATAAAGATATATCTGGAAGTATGATGATGATGATGATGATGATGATGATGATGAT 359
 Cp 348 ATAAAGATATATCTGGAAGTATGATGATGATGATGATGATGATGATGATGATGAT 407
 Db 360 ATGATTAACATCTAGCTGCTGCTGCTGCAATGACCAATGATGATGATGATGATGATGAA 419
 Cp 408 ATGATTAACATCTAGCTGCTGCTGCTGCAATGACCAATGATGATGATGATGATGATGAA 467
 Db 420 GCAAGTATGTCAGAGAGTGAATATGCGGTGAAGATTTAGCATATTTGCTATGCA 479

OY 468 GCAAGGATATGTCAGAGGTGAATTAATTCGGCTGAGATTCAGCATATTCGCTATGA 527

DB 480 TTGACGGAGGGGGAACCACTGGAAGTCATTCATTCATTAAGTGTGATGA 528

OY 528 TTGACGAGAGGGGAACCACTGGAAGTCATTCATTCATTAAGTGTGATGA 576

RESULT 13

LOCUS AA634065 572 bp mRNA EST 21-OCT-1997.

DEFINITION ac34c04.s1 Striatogene hnt neuron (#937233) Homo sapiens cDNA clone IMAGE:858342 3' similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ; mRNA sequence.

ACCESSION AA634065

NID 92557279

VERSION AA634065.1 GI:2557279

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 572)

AUTHORS Hallier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S., Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST project

JOURNAL Unpublished (1997)

COMMENT On Sep 19, 1997 this sequence version replaced gi:1520522.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 273.

FEATURES

source

1. 572

Location/Qualifiers

/organism="Homo sapiens"

/note="Vector: pBluescript SK-; Site-1: EcoRI; Site-2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Differentiated, post mitotic hnt neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATGGGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'."

/db_xref="taxon:9606"

/clone_1ib="IMAGE:858342"

/clone_1ib="Striatogene hnt neuron (#937233)"

/dev_stage="hnt neurons"

/lab_host="SOLR (kanamycin resistant)"

BASE COUNT 164 a 104 c 99 g 205 t

Query Match 40.7%; Score 518; DB 16; Length 572;

Best local Similarity 97.8%; Pred. No. 0.00e+00;

Matches 530; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

DB 31 CAGCAAAATATTTATTTATTCCTACATATGTAACATATATTCATATGTCCTCCCT 90

CP 1275 CAGCAAAATATTTATTTATTCCTACATATGTAACATATATTCATATGTCCTCCCT 1216

DB 91 TGCACATCTATTCACAGTACTTCGAATGACAACTGCTTTGATATTTAAGCATGTGCT 150

CP 1215 TGACACATCTATTCACAGTACTTCGAATGACAACTGCTTTGATATTTAAGCATGTGCT 1156

DB 151 AAAAGTATCTTGTGATGATGAAAGAAAGCTTTAATGGTAACTTCTGAGTATATT 210

CP 1155 AAAAGTATCTTGTGATGATGAAAGAAAGCTTTAATGGTAACTTCTGAGTATATT 1096

DB 211 GGATTAAGTACAGAGAGAAATTTACTTTAGTAGATGAGTCTTACAAATTTAAAGCTTGAA 270

CP 1095 GGATTAAGTACAGAGAGAAATTTACTTTAGTAGATGAGTCTTACAAATTTAAAGCTTGAA 1036

DB 271 AAGTACTACTTTTACTTTTATATATATCCAGATGACAGATGATAGCATATATGAGTGT 330

CP 1035 AAGTACTACTTTTACTTTTATATATATCCAGATGACAGATGATAGCATATATGAGTGT 976

DB 331 ATTCAGAGAAATCTCATTTAGTTTCTGATGATGAGAACACATTAATCCAGTCTGTGT 390

CP 975 ATTCAGAGAAATCTCATTTAGTTTCTGATGATGAGAACACATTAATCCAGTCTGTGT 916

DB 391 ACTGTGAGGAGAGATTCACAGGAGTGTGTTAAAGCATATCCAAATGGCTGTGAGCATCA 450

CP 915 ACTGTGAGGAGAGATTCACAGGAGTGTGTTAAAGCATATCCAAATGGCTGTGAGCATCA 856

DB 451 GGATCAGACTTGAAGGGGCTCTCAGACAAAGTGTATTCATGCAAGTATCTTTTCCA 510

CP 855 GGATCAGACTTGAAGGGGCTCTCAGACAAAGTGTATTCATGCAAGTATCTTTTCCA 796

DB 511 TTGGTTTCTTAAGTCTTATGCTTTTCCATGATGTCGCAAGTCTTTTAAATATCAAG 570

CP 795 TTGGTTTCTTAAGTCTTATGCTTTTCCATGATGTCGCAAGTCTTTTAAATATCAAG 736

DB 571 GC 572

CP 735 GC 734

RESULT 14

LOCUS AA702134 521 bp mRNA EST 19-DEC-1997

DEFINITION 2183g08.s1 Soares_fetal_liver_spleen_infls.s1 Homo sapiens cDNA clone IMAGE:447614 3' similar to SW:IPYR_BOVIN P37980 INORGANIC PYROPHOSPHATASE ; mRNA sequence.

ACCESSION AA702134

NID 92705247

VERSION AA702134.1 GI:2705247

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 521)

AUTHORS Hallier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S., Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST project

JOURNAL Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1394456.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 399.

FEATURES

source

1. 521

Location/Qualifiers

/organism="Homo sapiens"

/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site-1: Pac I; Site-2: Eco RI. This is a subcloned version of the original Soares fetal liver spleen INFs library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGAGAGAAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Felima Bonaldo."

/db_xref="GDB:1351871"
/map="22"
/clone_lib="Soares_fetal_liver_spleen_INTLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 156 a 100 c 96 g 169 t
ORIGIN

Query Match 40.3% Score 513; DB 17; Length 521;
Best Local Similarity 99.2% Pred. No. 0.00e+00;
Matches 517; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 CACCAAAATATTTTATTTCTTACATATGATATATCATATATGATGCTCCCT 60
|||
Cp 1275 CACCAAAATATTTTATTTCTTACATATGATATATCATATATGATGCTCCCT 1216
|||
Db 61 TGCACATCTATTCACAGTACTCTCCAAATGACAACTGCTTGTATTTAAGCATGCT 120
|||
Cp 1215 TGCACATCTATTCACAGTACTCTCCAAATGACAACTGCTTGTATTTAAGCATGCT 1156
|||
Db 121 AAAAGTTATCTTGTGATGATGAAAAATGCTTTAGATGATATATCTGATATATT 180
|||
Cp 1155 AAAAGTTATCTTGTGATGATGAAAAATGCTTTAGATGATATATCTGATATATT 1096
|||
Db 181 GGATTTAGTCACAGATATCTTGTGATGATGATGATGATGATGATGATGATGATGAT 240
|||
Cp 1095 GGATTTAGTCACAGATATCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1036
|||
Db 241 AACCTACTACTTTTACTTCTTATATACATCAGATGATGATGATGATGATGATGATGAT 300
|||
Cp 1035 AACCTACTACTTTTACTTCTTATATACATCAGATGATGATGATGATGATGATGATGAT 976
|||
Db 301 ATTCCAGAGAAATCTCAATGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 360
|||
Cp 975 ATTCCAGAGAAATCTCAATGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 916
|||
Db 361 ACTGTGAGAGCATATTCACAGGCTGTGTAAAGCATCCCAATGGCTGTGGAGCATCA 420
|||
Cp 915 ACTGTGAGAGCATATTCACAGGCTGTGTAAAGCATCCCAATGGCTGTGGAGCATCA 856
|||
Db 421 GGATCAGACTTGAAGGGGCTCTCAGACAAAGTTGTATTCATGCAATGATTCCTTTTCCA 480
|||
Cp 855 GGATCAGACTTGAAGGGGCTCTCAGACAAAGTTGTATTCATGCAATGATTCCTTTTCCA 796
|||
Db 481 TCCGTTGCTTACTACTAATGCTTTCCAAATGCTCAGAGAGT 521
|||
Cp 795 TCCGTTTCTTACTACTAATGCTTTCCAAATGCTCAGAGAGT 755
|||

RESULT 15
LOCUS AA075438 580 bp mRNA EST 23-DEC-1997
DEFINITION zmr7c06.r1 StrataGene ovarian cancer (#937219) Homo sapiens cDNA
clone IMAGE:544906 5' similar to SW:IPR_BOVIN P37980 INORGANIC
PYROPHOSPHATASE ; mRNA sequence.
ACCESSION AA075438
NID 91615309
VERSION AA075438.1 GI:1615309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 580)
AUTHORS Euleria; Primates; Catarrhini; Hominoidea; Homo.
Hillier, L., Clark, N., Dubuque, T., Eilston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck ESF Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:176447.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800;
Fax: 314 286 1810
Email: estewarton.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 419.

FEATURES

1. 580
Location/Qualifiers
/organism="Homo sapiens"
/note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Papillary serous carcinoma, isolated from ascites, 64 year
old Caucasian. Average insert size: 0.8 kb; Uni-ZAP XR
vector; -5' adaptor sequence: 5' GAATTCGGGCGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'
/db_xref="GDB:3923094"
/db_xref="taxon:9606"
/clone="IMAGE:544906"
/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT 181 a 107 c 141 g 149 t 2 others
ORIGIN

Query Match 40.3% Score 514; DB 36; Length 580;
Best Local Similarity 97.4% Pred. No. 0.00e+00;
Matches 557; Conservative 0; Mismatches 9; Indels 6; Gaps 6;

Db 9 GACAATGACCATTATGATGCTGTGAATGGAAGCAAGTATGCAAGAGTGAATA 68
|||
Oy 434 GACAATGACCATTATGATGCTGTGAATGGAAGCAAGTATGCAAGAGTGAATA 493
|||
Db 69 ATTGGCGTGAAGTTCTAGCATATTTGGTATGATGACGAAGGGGAACCGACTGGAAA 128
|||
Oy 494 ATTGGCGTGAAGTTCTAGCATATTTGGTATGATGACGAAGGGGAACCGACTGGAAA 553
|||
Db 129 GTCAATGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 188
|||
Oy 554 GTCAATGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 613
|||
Db 189 AAAGGCTGAACCTGGCTCTTGAAGCTACTGTGTGATGATGATGATGATGATGATGATGAT 248
|||
Oy 614 AAAGGCTGAACCTGGCTCTTGAAGCTACTGTGTGATGATGATGATGATGATGATGATGAT 673
|||
Db 249 CCGATGGAAGAAACCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 308
|||
Oy 674 CCGATGGAAGAAACCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 733
|||
Db 309 GCCATTGATATTTAATAACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 368
|||
Oy 734 GCCATTGATATTTAATAACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
|||
Db 369 CCGAATGGAAGAAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 428
|||
Oy 792 C-GAATGGAAGAAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 850
|||
Db 429 GATCCTGATCTGCGAGAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
|||
Oy 851 GATCCTGATCTGCGAGAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 910
|||

Db 489 ACCAGTACCCAGACAGCTGGATTAAGTNGTTCATCATCACCGAAACCTNATGGAGATTCC 548
|||
QY 911 AC-AGTACCAACAGACAGCTGGATTAAGTNGTTCATCATCACCGAAACCTAATG-AGATTCT 968
|||
Db 549 CTGGGATTCCAGCTGATTTATTCACCATGCTG 580
|||
QY 969 CTGGAATACAGC-TGATTTCTCTCATCTG 999
|||

Search completed: Mon Dec 27 12:53:35 1999
Job time : 4587 secs.

RESULT	3	CEC47E12	37855 bp	DNA	INV	23-NOV-1998
LOCUS		Caenorhabditis elegans	cosmid C47E12,	complete sequence.		
DEFINITION		268882				
ACCESSION		268882				
NID		g1167468				
VERSION		268882.1	GI:1167468			
KEYWORDS		HNG: ADP/ATP carrier protein; Alpha-mannosidase; Inorganic pyrophosphatase; Seryl-tRNA synthetase; Ubiquitin-activating enzyme; Ulp1 like; Yeast YD78 like.				
SOURCE		Caenorhabditis elegans.				
ORGANISM		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Pelodietinae; Caenorhabditis				
REFERENCE		1 (bases 1 to 37855)				
AUTHORS		Coles, L.				
TITLE		Direct Submission				
JOURNAL		Submitted (25-JAN-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu				
REFERENCE		2 (bases 1 to 37855)				
AUTHORS		Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copey, T., Cooper, J., Coulson, A., Craton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawks, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N., Lattelle, P., Lightning, J., Lloyd, C., McMuray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, R., Waterston, R., Watson, A., Wellstock, L., Wilkinson-Sproat, J. and Wohlman, P.				
TITLE		2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans				
JOURNAL		Nature 368 (6466), 32-38 (1994)				
MEDLINE		94150718				
COMMENT		Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis, see: - http://webace.sanger.ac.uk/cgi-bin/displayDb-wormaceclass-sequence/object-C47E12 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. IMPORTANT: This sequence is not the entire insert of clone C47E12. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone C47E12 is at 1 in this sequence. The true left end of clone F44D12 is at 37752 in this sequence. The true right end of clone T13HD10 is at 4978 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 269361. The end of this sequence (37752..37855) overlaps with the start of sequence 268298. Location/Qualifiers 1..37855 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="IV" /clone="C47E12" 959..2262 /gene="C47E12.11" join(959..1428,1483..1578,1623..1728,1787..2112,2160..2262) /gene="C47E12.11"				

gene
/codon_start=1
/protein_id="CA93109.1"
/db_xref="PID:e1344633"
/db_xref="PID:g3875114"
/db_xref="GI:3875114"
/db_xref="SPTREMBL:Q18682"
/translation="MIRARSMETLRDHRINGLTIYKRRREDDVEKKMVRVTLPEADEN
SIDSCVSKSLPPRSASNSKROASGASTASKNKTSEKSKTSGRCHVAKKWKRRK
KTKSESRERLDDVTIQTLDLTKLNTSTIRFPHIGTCCPHLELTKSKRADLOSSK
DSRSFKTGGELVNLNLAHSVKEYDEKGISMINVAQTPEPNACVSTKRRARAKMA
LAKGMSNDIASSTSSKSHKTKNTSDSYSDMSPVFLRVSSDTPEPGATSPATEE
RVRLMKLYALAREHIKDSYDYGSHSKNSKSKKNTQDTSKTSKSLSKMTYSESS
QLRPSDSCSEDSMDKOKSKKDVAKKKKGTROM"
3910..5046
/gene="C47E12.10"
join(3910..3969,4027..4230,4280..4400,4444..4622,
4669..4767,4819..5046)
/gene="C47E12.10"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA93102.1"
/db_xref="PID:e1344626"
/db_xref="PID:g3875107"
/db_xref="GI:3875107"
/db_xref="SPTREMBL:Q18675"
/translation="MVLTVLITKSLVLLVGLTPEPPYIYIIILSTNLFLOGSKKK
KEKSESEDETEDESENORRRKRAADKKRKKITAAVAOAKKCGPVPRPKRAAI
AETIDPNQTYLAGLNLVDLEQEPGAPRAPPSNGMGTADPNQTYLAGLNLNDL
FEGGGGGAAPKAPATVNGSKAATNDPNQTYLADLNEIEFEKAKPEENAKAATNDP
NYQTLAGLNDEIEFKKPPAPVPPARVDQTPFINNDLRLPKQPVOKNVKAGYDPR
YOTLAGLNDEIFGADKKKNY"
5765..6522
/gene="C47E12.9"
join(5765..5839,5951..6028,6074..6265,6313..6522)
/gene="C47E12.9"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA93103.1"
/db_xref="PID:e1344627"
/db_xref="PID:g3875108"
/db_xref="GI:3875108"
/db_xref="SPTREMBL:Q18676"
/translation="MANICLSEITVLVLSFFKNVCEDLTIFFNEONNEDSKTYM
KATIDNNAVPAIIPCAVGRKRDVLANPKEVANEVFNVDALKTISDKVLEKS
ISLGFEESELEKIRELFEWEVEMRSSQTVGFKNFQCFKRIKENYSSEVAOKFEDL
SEEVESDFEKLQHLNENELIKE"
9248..11771
/gene="C47E12.8"
join(9248..9333,9383..9563,9608..9779,9828..10084,
10131..10363,11090..11402,11451..11771)
/gene="C47E12.8"
/note="predicted using GeneFinder; Similarly to Mouse
Ulip protein (PIR Acc. No. S55525,"
/codon_start=1
/protein_id="CA93104.1"
/db_xref="PID:e1344628"
/db_xref="PID:g3875109"
/db_xref="GI:3875109"
/db_xref="SMTS-PROT:Q18677"
/translation="MSLITKGTITVNDIAFKSDVILVDGRIEVIAPISIOPTPELVV
DATDRVLVPGSIDPHTHQQLPFMGEIAKDDPHRTTEAVAGGTMIIDYPIPTGSEL
LVADRMWGAADPVVCDYGLSMAITSGSPRIAEMLVGAEGINSKFPFLAVAYAG
FLVADREYQCMICAKIARAVHAEGSVIABRCHLSSGITGEGHQSRPEL
EAEAFRCATASOANCLVYVWVHSGAALAHNHRKGAIVGGEPIAGLADGSH
YVNEHMLAAVYVWSPILSRPSTPSALMKTLAGLGLHLTRATDCKTFCDOOKSLGKDD
FTKIPNGVNGVEDKRSVYMDGYHAGITIDPRFPAVYSTAAKIFNCPQKRIAVSS
DADIVYMANNAKRTISKOTHHAIIDFNIFBGMVGHVPELTISGRVYANNGOLKTVQ
GSGGFIPLAPDSQIVFSAVNDRRKRAPEVKIDRIPEPSALQTPDANANIVYAPRA
AIPGGASSIOF"
Complement(11905..13328)
/gene="C47E12.7"
Complement(join(11905..12121,12209..12722,12773..12884,
12930..13153,13202..13328))

Db	Query Match	Score	DB	Length	873:
Best Local Similarity	63.8%	Pred. No. 1.44e-68;			
Matches	329;	Conservative	0;	Mismatches 187;	Indels 0; Gaps 0;

140 ACATGTTGGTGGAGGCTCCACGCTTTGACGACCAACGGAAGAAAGGAGATTCCTTACCTTAAATTCACACACA

510 AGAAGCTTACGCAATATTTTACACCTCTTGACATACATACCTTGCTCCAAATTCACACACA 451

27375 TCAATTTGATCATTTGCTCTCTTTAGCTCCAGATGTACAGAACACATGGTTTGATCTTCC

Db 27375 TCAATTTGATCATTTGCTCTCTTTAGCTCCAGATGTACAGAACACATGGTTTGATCTTCC 27434

450 TCAATTTGGGATCATTTGTCCACACACACACACAGATGTATTTATCATTTGTGCCCTGGGCTTCC

CP 450 TCAATTTGGGATCATTTGTCCACACACACACAGATGTATTTATCATTTGTGCCCTGGGCTTCC 391

27435 CATCTCTGTGGAAAGCGCGTATCTTCACATGATATCTTGTGTGGGAAATTTGTGA

Db 27435 CATCTCTGTGGAAAGCGCGTATCTTCACATGATATCTTGTGTGGGAAATTTGTGA 27494

390 CAATCTCTAGAGGATGACACCATATCTTCAGATATATCTTTTATACGGGAACAAATTTGCA

CP 390 CAATCTCTAGAGGATGACACCATATCTTCAGATATATCTTTTATACGGGAACAAATTTGCA 331

27495 ACAAGCGAGCTACACCCCTTCTCTCATCTTGTGTTGATGGCAGATGCTTCCTGTG

Db 27495 ACAAGCGAGCTACACCCCTTCTCTCATCTTGTGTTGATGGCAGATGCTTCCTGTG 27554

330 ACAATGCGAGATTTTCTTTTTCACATCTTGTTTAATAGGTTTAAAGGTTTCCCTTTGTA

CP 330 ACAATGCGAGATTTTCTTTTTCACATCTTGTTTAATAGGTTTAAAGGTTTCCCTTTGTA 271

27555 GCCATTTTCATTTTAGATTTGTCCACACGCGAATCTCAACATCATGT

Db 27555 GCCATTTTCATTTTAGATTTGTCCACACGCGAATCTCAACATCATGT 27603

270 GCATCTTCATTTTGTGATTTAGACACGCGTGATCTCAACTACATCATGT

CP 270 GCATCTTCATTTTGTGATTTAGACACGCGTGATCTCAACTACATCATGT 222

4 AF085601 873 bp mRNA INV 18-DEC-1998

RESULT LOCUS DEFINITION Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38

AF085601 mRNA, complete cds.

ACCESSION AF085601

3600093

NID 93600093

AF085601.1 GI:3600093

VERSION KEYWORDS

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea;

Drosophilidae; Drosophila.

1 (bases 1 to 873)

Gdula,D.A., Sandlitopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.

REFERENCE AUTHORS TITLE

inorganic pyrophosphatase is a component of the Drosophila

nucleosome remodeling factor complex

Genes Dev. 12 (20), 3206-3216 (1998)

JOURNAL MEDLINE 99003073

2 (bases 1 to 873)

Gdula,D.A., Sandlitopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.

Direct Submission

FEATURES

Submitted (20-AUG-1998) LMCB, National Cancer Institute, NIH,

Building 37, Room 5E-20, Bethesda, MD 20892-4295, USA

Location/Qualifiers

1..873

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2"

/map="60C"

1..873

/gene="Nurf-38"

1..873

/note="nucleosome remodeling factor complex component"

/codon_start=1

/product="inorganic pyrophosphatase NURF-38"

/protein_id="AAC97112.1"

/db_xref="PIR:G3600094"

/db_xref="GI:3600094"

/translation="MALTYEVEGAKNSPSLYFKKNCNVISPMHIDILYANEKTT

ITYNNVVEPRYTNAKMELISLTKENPNLKODIKKGKLEFVANCPEPKGYINNGALPDT

WENPDHEPSTGCKGNDPLIDVLEIGYRAKRGDVLKVLGQFALIDEGEETPKIILA

IDVNDPLASKVNDIADVDQYFPGLLRATVEMFKYIKIPDGCKPEQFAFNDAKNADPA

NTIIAEIKFQNDILVHOSPASGISISTNINNRNSEHPIPREAEKIIIAEPDQGOVE

VSDLPDWHETHLK"

BASE COUNT 228 a 247 c 233 g 165 t

ORIGIN

Query Match

Best Local Similarity

Score 142; DB 22; Length 873;

Matches

Pred. No. 1.44e-68;

329;

Conservative

0;

Mismatches 187;

Indels

0; Gaps 0;

```

promoter      128. 131
              /note="pot. TATA box"
CDS           142. 552
              /note="Histone H3 (AA 1-136)"
              /codon_start=1
              /protein_id="CAA32444.1"
              /db_xref="PID:g2901"
              /db_xref="GI:2901"
              /db_xref="SWISS-PROT:P02303"
              /translation="MARTKOTARKSTGKAPRKQLASKAARKSNPSTGYKKRHPKP
              GTVALREIRREKSTELLIRLPEQLVREIAQCFKDLRFQSSAIGALQESVEAYLV
              SLFEDTNLAHAHKRVTIQKKDKILARLNGERS"
misc_feature  580. 621
              /note="polypyrimidine tract"
              664. 667
              /note="pot. TATA box"
              710. 713
              /note="pot. TATA box"
              742. 745
              /note="pot. TATA box"
              765. 768
              /note="pot. TATA box"
              866. 1498
              /note="N-(5'-phosphoribosyl)anthranilate isomerase (AA
              1-210)"
              /codon_start=1
              /protein_id="CAA32445.1"
              /db_xref="PID:g2902"
              /db_xref="GI:2902"
              /db_xref="SWISS-PROT:P13997"
              /translation="MLVAVCGLOQVFAAKTAVDDGADYLCITICVPGKKRTIDSSVAKG
              ISTVAHOQENLKVGVFERNQSVDDVLQLYHEYNNLDVQLMGDEDDIKERYSLRIPSS
              IPIKRRQPDCCEDLLDILYEHVDNVLTFPDSGSGGTEKLMNSAISMGSASPEIKF
              IIAAGLNDGNSVAINMLPNAIGVDSGGVETDGIKDLKRYKLETOAAS"

```

```

misc_feature
/note="TPP open reading frame before probable AUG start
codon"
complement(2453..2460)
/note="pot. TATA box"
complement(2475..2480)
/note="pot. TATA box"
complement(2486..2490)
/note="pot. TATA box"
complement(2513..2518)
/note="pot. TATA box"
complement(2524..2799)
/note="unidentified open reading frame (AA 1-91)"
/codon_start=1
/protein_id="CAA32447.1"
/db_xref="PID:g2904"
/db_xref="GI:2904"
/db_xref="SWISS-PROT:P13998"
/translation="MSYTRPVGAKNSLDYKVIIEKDGKPIISAFHDIPLIADENANGIF
NMVEIPIRWTNAKLEITKEEPLNPIIOTDRKKGLRYRNCFPPIHGITHNGATPOTWE
DPNESHPEITKAVDNDPLDYLEIGEOVAAYGVQKAVLGMALDLEGEDVIALD
INPLAPKNDIEDVEKHLPEGLRATVEMERIKIIPDGKEPNOFAFSGEAKNKRYTLD
VIECEENAMKLLISGSADAKIDLNTITSDTATYSAAEASAVPANVLPDEPIDRS
IDWFFISGSA"
complement(2442..2561)
/note="TPP open reading frame before probable AUG start
codon"
complement(2453..2460)
/note="pot. TATA box"
complement(2475..2480)
/note="pot. TATA box"
complement(2486..2490)
/note="pot. TATA box"
complement(2513..2518)
/note="pot. TATA box"
complement(2524..2799)
/note="unidentified open reading frame (AA 1-91)"
/codon_start=1
/protein_id="CAA32447.1"
/db_xref="PID:g2904"
/db_xref="GI:2904"
/db_xref="SWISS-PROT:P13998"
/translation="MIRPNDFKNTLCNCNCIOSDOIYTDPNVYPSPTPIHHRISFHL
NYRFFKQASFPSCRLKIVDDLRKKKLQNSYINRNEKCHWFTIGH"
complement(2569..2574)
/note="pot. TATA box"

BASE COUNT      832 a      550 c      599 g      906 t

ORIGIN
Query Match
-- 10.8% Score 137 DB 27 Length 2887

```


Note: remainder of annotations omitted.

```
Query Match          10.8%; Score 138; DB 29; Length 184180;
Best Local Similarity 78.5%; Pred. No. 2,98e-85;
Matches    204; Conservative   0; Mismatches 54; Indels   2; Gaps   2;
```

Db 135720 AAGGATCATGTTGTGTGAATAACACAGATCTCAGAGCCCT -CAAGTGTATCTTCAC 135778
|||||
QY 800 AAAGAATCGATTGCATGAAATACAACTTTGTGTGAGGCCCTTAAGTGATCTTAT 859

Db 135779 ATAGCTAACCATTGTGGATGCTTCCACCAACATGTATAGCTTATGTAGCACG 135838
||
QY 860 GCCTCCAGAGACCAATGTGTGATGCTTTACCAACACCACTGTGAAATCTGCTGCAC -AGTACC 918

Db 135839 CACGATGTAGAGAGCTGTGTCCAACACCAAGAAAAAGAAATAAAACCTTCTTGATACG 135898
|||||
QY 919 AACGAGAGCTGATAGTAGGTTCCATACACAGAAAAACTATGAGATTTCTGTGANAACA 978

Db 135899 AGCTGATATTGCTGTCTGTATTATTCATCATCTGGAAGTATGAGATCAAAGTAGCAGCTTTTC 135958
|||||
QY 979 AGCGATATTGCTACATCGTTCATCTGAGATGTATGAAATAAAGATAGACCTTTTC 1038

Db 135959 AAGGCTTAGGCTTTAGAA 135978
|||||
QY 1039 AAGGCTTTAATTTGTAGAA 1058

RESULT 7
LOCUS SCPPAG 1612 bp DNA PLN 12-SEP-1993
DEFINITION Submitted PPA gene for Inorganic Pyrophosphatase (EC 3.6.1.1).
ACCESSION X13253
VERSION 94198
KEYWORDS X13253.1 GI:4198
 inorganic pyrophosphatase; phosphatase; PPA gene; pyrophosphate
 phosphorylase.
SOURCE Baker's yeast.
ORGANISM Saccharomyces cerevisiae
COMMENTARY Saccharomycetes; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Eukaryota; Fungi; Ascomycota; Department of Chemistry, 231 South 34th Street,
 Philadelphia, PA 19104
REFERENCE 1 (bases 1 to 1612)
TITLE Kolakowski, L.F.
AUTHORS Direct Submission
PUBLISHED Submitted (14-OCT-1988) Kolakowski L.F., University of
 Pennsylvania, Department of Chemistry, 231 South 34th Street,
 Philadelphia, PA 19104
REFERENCE 2 (bases 1 to 1612)
TITLE Kolakowski, L.F., Jr., Schlosser, M. and Cooperman, B.S.
AUTHORS Cloning, molecular characterization and chromosome localization of the
 inorganic pyrophosphatase (PPA) gene from *S. cerevisiae*
 Nucleic Acids Res. 16 (12):10441-10452 (1988)

MEDLINE 89083474
 FEATURES
 source location/Qualifiers
 1..1612
 /organism="Saccharomyces cerevisiae"
 /strain="X2180"
 /db_xref="taxon:4932"
 /chromosome="chromosome 2"
 /clone.lib="lambda gt10.1"
 /clone="PM14"
 1..1612
 /gene="IPP1"
 /db_xref="SGD:L0000872"
 4..9
 /gene="IPP1"
 /note="TATA-like box"
 /db_xref="SGD:L0000872"
 69..73
 /gene="IPP1"
 /note="pot. transcription start site"
 /db_xref="SGD:L0000872"
 109..117
 /gene="IPP1"
 /note="TATA-like box"
 /db_xref="SGD:L0000872"
 139..143
 /gene="IPP1"
 /note="pot. alt. transcription start site"
 /db_xref="SGD:L0000872"
 216..220
 /gene="IPP1"
 /note="pot. alt. transcription start site"
 /db_xref="SGD:L0000872"
 227
 /gene="IPP1"
 /note="pot. alt. transcription start site"
 /db_xref="SGD:L0000872"
 233..249
 /gene="IPP1"
 /note="constitutive promoter"
 /db_xref="SGD:L0000872"
 255..260
 /gene="IPP1"
 /note="TATA-like box"
 /db_xref="SGD:L0000872"
 291..297
 /gene="IPP1"
 /note="TATA-like box"
 /db_xref="SGD:L0000872"
 361
 /gene="IPP1"
 /note="pot. alt. transcription start site"
 /db_xref="SGD:L0000872"
 390..1253
 /gene="IPP1"
 /note="pyrophosphatase (AA 1 - 287)"
 /codon.start=1
 /db_xref="SGD:L0000872"
 /protein_id="CAA31629.1"
 /db_xref="PIR:g4199"
 /db_xref="GI:4199"
 /db_xref="SWISS-PROT:P00817"
 /translation="MTYTRIGAKNTLEKVIYIKGKRPVSAFHDIPIYADKENNIF
 NMVEIPRTNARKLEIKETLNPIIDDTCKGLRFRNCFPHGTHINAGAPQWTE
 DPVSHRETKAGNDNPIDLEIGETIATGQYKQVKAUGIMALLDEGETDMKVIAD
 INPLAKLNDIDVEKTFPGGLRAINEMFRITKIDGKRNQFAFSGEKKKKVIALD
 IIEKHSWMQLAGKSSDKGIDLNVTLDPDPTVSKAASDAIPPASLEKADAPIDKS
 IDKWFITSGSV"
 BASE COUNT 496 a 321 c 306 g 489 t
 ORIGIN
 Query Match 10.3% Score 131; DB 27; Length 1612;
 Best Local Similarity 61.8%; Pred. No. 1,82e-79;
 Matches 344; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

Db 533 ACATGTTGTTGAATTCACAGTGGACCAAGCCAGTGAATTCACCAAGAA 582
 Oy 222 ACATGTTGTTGAATTCACAGTGGACCAAGCCAGTGAATTCACCAAGAA 281
 Db 563 CTTTAACCAATTCATCCCAAGACCAAGAGGCAAGTGAATTTGTAAGATCTG 642
 Oy 282 CTTTAACCAATTCATCCCAAGACCAAGAGGCAAGTGAATTTGTAAGATCTG 341
 Db 643 TCCCTCACCATTGTTACATTCACCAATTCAGTGGCTTTCCCAAACTTGGAAACCCAA 702
 Oy 342 TCCCTCACCATTGTTACATTCACCAATTCAGTGGCTTTCCCAAACTTGGAAACCCAA 401
 Db 703 AGTAAAGCCACCAAGAACTAAGCAGTGGTGAACCAAGATTCATGATGTTGAAA 762
 Oy 402 GGCACAAATGATTAACATACCTGCTGTTGTGTGACAAATGACCAATGATGTTGAAA 461
 Db 763 TTGTTGAATCTATTGCTTACCTGCTCAAGTCAAGCAAGTTAAGGCTTACGATTCATG 822
 Oy 462 TTGTTGAATCTATTGCTTACCTGCTCAAGTCAAGCAAGTTAAGGCTTACGATTCATG 521
 Db 823 CTTTATGATGATGAGGAGAGACGATTCGAAGTGAATTCGATGATTAACATTCAT 882
 Oy 522 CTTTATGATGATGAGGAGAGACGATTCGAAGTGAATTCGATGATTAACATTCAT 581
 Db 883 TAGCCCAAAATTAAGACGACATTAAGATGTTGAATTAATCTCCAGGCTATTGAGG 942
 Oy 582 ATGACGACCAATTAATTAAGATTAATGATGATGATTAATCTCCAGGCTATTGAGG 641
 Db 943 CTACTAAGCAATGTTCAAGATTTTCAAAATCCCAAGTGAAGCCAAATTCAT 1002
 Oy 642 CTACTAAGCAATGTTCAAGATTTTCAAAATCCCAAGTGAAGCCAAATTCAT 701
 Db 1003 CCTTCCTCCGAGTGAAGCAAGCAAGTGAAGTGAATTCATTAACCAAGCAATG 1062
 Oy 702 CGTTTATGAGCAATTTAAGATTAAGCAATTCGATGATTAATTAAGCAATTCAT 761
 Db 1063 ACTCTCGAAGCAATTA 1079
 Oy 762 ACCATGGAAGCAATTA 778

RESULT 8
 LOCUS SCYBR011C 2411 bp DNA 11-AUG-1997
 DEFINITION S.cerevisiae chromosome II reading frame ORF YBR011C.
 ACCESSION Z35880 Y13134
 NID 9536205
 VERSION Z35880.1 GI:536205
 KEYWORDS
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomyces.
 REFERENCE
 AUTHORS Entlian,K.D., Koetter,P., Rose,M., Li,Z., Therman,R., Brendel,M.,
 Baur,A., Boles,E., Miosga,T., Schaaff-Gerstenschlaeger,I. and
 Zimmermann,F.K.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2411)
 TITLE MIPS
 JOURNAL Direct Submission
 SUBMITTED (30-AUG-1994) Data collected by MIPS on behalf of the
 European yeast chromosome II sequencing project. MIPS at the
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
 Martinsried, FRG; E-mail: Mews@mips.emblnet.org
 3 (bases 1 to 2411)
 REFERENCE
 AUTHORS Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Bactlet,M.C.,
 Barthe,C., Baur,A., Becam,A.M., Billeau,N., Boles,E., Brandt,T.,
 Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C.,
 Contreras,R., Crouzet,M., Czapluch,C., Demolis,N., Delaveau,T.,
 Dignonon,F., Domdey,H., Duesterhus,S., Dubois,E., Dujon,B., El
 Bakoury,M., Entlian,K.D., Feuerhahn,M., Fiers,W., Fobio,G.M.,
 Fritz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A.,

gene complement(1110..1967)
 CDS /gene="IP1"
 /complement(1110..1967)
 /gene="IP1"
 /EC_number="3.6.1.1"
 /codon_start=1
 /product="inorganic pyrophosphatase (pyrophosphate phospho-hydrolase)"
 /protein_id="CA04453.1"
 /db_xref="PID:el180018"
 /db_xref="PID:g2624379"
 /db_xref="GI:2624379"
 /db_xref="SWISS-PROT:013505"
 /translation="MSYTRQGAATLNNRFFIEKDGQVSPFHDIPLYADESKVLNVAVEPRTNAKLEISKKEKLNPILODRKGLRFVNRNCFPHGFIHNYGAFPOIWE DNPVHPETKAGDNDPLDVCIEIGERSYGOYVLEVMALLDEGETDMKVIYDI NDPLAKNDIEDVEKHMPLRATNEMERIKIPGCKENOFASFGECKNRYAEV IOEERAMEKLIIEGTSPEIDJLNTLTOSTSFSPSATSNAPTPAPAKIDSID KMFYISGS"

BASE COUNT 661 a 488 c 540 g 636 t 1 others

Query Match 9.9%; Score 126; DB 27; Length 2326;
 Best Local Similarity 63.4%; Pred. No. 2,38e-75;
 Matches 334; Conservative 0; Mismatches 190; Indels 3; Gaps 3;

Db 1324 GCATCTCTGTTGTTGATTCACAGAAAGCAAACTGTTCTCAGGTTTCCATCA 1383
 Cp 735 GCMAAGCTCTATCTTAAATCTGATTAAGCAAACTGTTCTGTTTCCATCA 676
 Db 1384 GGGATCTTGAATTTGAAACCACTGTTGCTCTCAACAAGCCCGGATGCTTT 1443
 Cp 675 GGAACCTTATACCTTTAAACCAAGCTGCTTCAAGTACGAGGTTTCAACCGT 616
 Db 1444 TCACATCTCGATGTCATGAGTACGAGCAATAGGCTGTGATGTCAGAGATG 1503
 Cp 615 TTGACATCATGATATCATATTAATGGCTGATCAGATCATCCATTAATGCGCAT 556
 Db 1504 ACCTTCAGTACGCTCTCTCCATCCAAAGCCATAACCGAGGACTTGAATTGC 1563
 Cp 555 ACTTCCAGTCTGCTTCCCTTGTCAATCATAGCCAAATGCTTGAATTCACGCCA 496
 Db 1564 TTGACCTGTCAGTATAC-TAGCT-CA-CCAATTCGAAACGTCATGATGAT 1620
 Cp 495 ATATATTCACCTCTTCAACATCTGCTTCAATTCACACATCAATTTGGTAT 436
 Db 1621 TCACCTTGCGCTTGTCTCTGATGAGTACGTTAGGCTTCCCATGCTGAGGAG 1680
 Cp 435 TCACCAACAGCAAGCATATGTTATCATCTGCGCTTCCCAAGTCTGAGGAG 376
 Db 1681 GCACGCTAGTGTGAATGATACCGTGTGAGGAAACAGTTCAGAACGAAAGTTT 1740
 Cp 375 GCACCATAGTCTCAGATATATCTTATACGGGAAACAAATTCACCATAGCAAGTTT 316
 Db 1741 CCCTTCTTGTGCTCGAAGATGAGTTCATTTCTCTCTGGAATCTCAACTTG 1800
 Cp 315 CCTTTTACATCTGTTTAAAGGTTTAAAGGCTTGTGACCAATCTCCATTTT 256
 Db 1801 GGCTTGTCCAGGGGAGCTCAACACCATGTTGAAACCTTCT 1847
 Cp 255 GCATTAGACCAAGCGTGTCTTCACTACCATGTGAACACATCTT 209

LOCUS 10 SPPAG 1366 bp DNA PLN 12-SEP-1993
 DEFINITION Yeast ppa gene for inorganic pyrophosphatase (EC 3.6.1.1).
 ACCESSION X54301
 VERSION 95013
 X54301.1 GI:5013
 KEYWORDS inorganic pyrophosphatase; ppa gene; pyrophosphatase; pyrophosphate phosphohydrolase.
 SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe
 Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomyces
 1 (bases 1 to 1366).
 REFERENCE
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (08-AUG-1990) Kawasaki I., The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai Minato ku, Tokyo 108, Japan
 2 (bases 1 to 1366)
 REFERENCE
 AUTHORS
 TITLE Nucleotide sequence of S. pombe inorganic pyrophosphatase
 JOURNAL Nucleic Acids Res. 18 (19), 5888 (1990)
 MEDLINE
 COMMENT
 FEATURES
 source
 location/Qualifiers
 1..1366
 /organism="Schizosaccharomyces pombe"
 /db_xref="taxon:4896"
 42..50
 /note="put. TATA box"
 90..94
 /note="put. mRNA start sequence"
 151..1020
 /note="inorganic pyrophosphatase"
 /codon_start=1
 /protein_id="CA38199.1"
 /db_xref="PID:g5014"
 /db_xref="GI:5014"
 /db_xref="SWISS-PROT:P19117"
 /translation="MSEYTRQGAATLNNRFFIEKDGQVSPFHDIPLYADESKVLNVAVEPRTNAKLEISKKEKLNPILODRKGLRFVNRNCFPHGFIHNYGAFPOIWE DNPVHPETKAGDNDPLDVCIEIGERSYGOYVLEVMALLDEGETDMKVIYDI NDPLAKNDIEDVEKHMPLRATNEMERIKIPGCKENOFASFGECKNRYAEV IOEERAMEKLIIEGTSPEIDJLNTLTOSTSFSPSATSNAPTPAPAKIDSID KMFYISGSPL"
 misc_feature
 1142..1147
 /note="put. polyA signal"
 BASE COUNT 404 a 297 c 258 g 407 t

Query Match 7.8%; Score 100; DB 27; Length 1366;
 Best Local Similarity 59.7%; Pred. No. 3.19e-54;
 Matches 308; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Db 266 ATGCTGAGAGCAACCTTGAACATGATGCTGAGATCCCTGTTGATCAGGCTAAC 325
 Cp 201 ATGCAATAGAGATGTTTCAATGATGATGAGTACAGCTGCTGATATGCAAAA 260
 Db 326 TTGAATTAACAAGAGGCTTGAACCTTGAACCTTCAAGCAAGATTAAGAAGTTAGC 385
 Cp 261 TGAAGATGCTCAAAAGGACCTTTAAACCTATTAACCAAGATGGAAGAAAAAC 320
 Db 386 TTGCTTCTGTTGTAACGCTTCTTCCACACGCTTACATTTGGAACACTAGCTTTTC 445
 Cp 321 TTGCTATGTTGGAATTTGTTCCGTATTAAGGATATCTGGAACATATGCTGCATCC 380
 Db 446 CTCAAACCTATGAGAGTCTATGTTGTTCACTCCTGAACCAAGGCCAAGCTGATAGC 505
 Cp 381 CTCAGACTTGGGAAGCCAGGCGCAATGATTAACATATCTGCTGTTGTGTCACATG 440
 Db 506 ATCTCTCATGTTTGTGATGATGAGTGAAGCTCGTGCTACACGCCCAAGTCAAGCAAG 565
 Cp 441 ACCCAATGATGTTGTGTAATTTGAAGAGAGATGATGCAAGAGAGTGAATATTTGGCG 500
 Db 566 TTAAGCTCTGGGTGTATAGCCCTTTTGAGAGAGGTGAGACTGAGTGAAGTTATTTG 625
 Cp 501 TGAAGTTCTAGCATATGCTATGATGAGAGAGGGAACCAAGTGAAGTGAAGTATG 560
 Db 626 TCATGAGCTTAATACCTTGAAGCTCAACCTTAATGATTTTGAAGATGTTGAGATGTC 685
 Cp 561 CCATTATGATGATATCTGATGACCAACATTAATGATATCAATGTCACAAAGCGC 620

Db 686 ATATGCTGCTTATCCGTCCTACTACAGTGTCCGATTACACAGATCCCGATG 745
 OY 621 TGAATCTGCTACTAGAACCTACTGTGACTGTTAGAGGTATTAAGGTTCTCATG 680
 Db 746 GCAGCCTGAAATATCTTCTGCTTCACTGATGTGAT 781
 OY 681 GAAACACGAAATAGAGTTTCTGTTATGCAAGAT 716

RESULT 11
 LOCUS SPAC23C11 38391 bp DNA PLN 17-JUL-1998
 DEFINITION S.pombe chromosome I cosmid c23C11.
 ACCESSION 298559
 MID 92330764
 VERSION 1 GI:2330764
 KEYWORDS 405 ribosomal protein; alanyl-tRNA synthetase; casein kinase II;
 cal; dna mismatch repair; Homol D box; Homol E box; inorganic
 pyrophosphatase; mitochondrial dna repair; muts family; paired
 amphipathic helix; php3; ploi; polo box; ppai; ribosomal protein
 s12; ribosomal protein s23; rps23; serine/threonine-protein kinase;
 transcriptional activator; zinc binding; zinc cadmium resistance
 protein; zinc metalloproteinases; zinc/cadmium resistance.
 fission yeast.
 Schizosaccharomyces pombe
 Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomyces.
 1 (bases 1 to 38391)
 Brown, D. and Churcher, C.M.
 Unpublished
 2 (bases 1 to 38391)
 Barrell, B.G., Rajandream, M.A. and Wood, V.
 Direct Submission
 Submitted (13-AUG-1995) Schizosaccharomyces pombe chromosome I
 sequencing project. Sanger Centre, Hinxton Hall, Hinxton, Cambridge
 CB10 1RQ E-mail: barrell@sanger.ac.uk

SOURCE
 ORGANISM
 Schizosaccharomyces pombe

REFERENCE
 AUTHORS
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 JOURNAL

COMMENT
 Notes:
 Details of yeast sequencing at the Sanger Centre are available on
 the World Wide Web.
 (URL, <http://www.sanger.ac.uk/yeast/home.html>)
 Protein coding regions (CDS) have been predicted with the help of
 computer analysing the GeneFinder program in PomBase (an ACEDB
 database) with additional predictions for the branch-acceptor sites
 supplied by the program Sp3splice. CAUTION: It is possible that for
 any individual CDS we may have underestimated or overestimated the
 number of introns/exons or we may not have chosen the correct
 splice donor/acceptor sites. CDS are numbered using the following
 system eg SPAC3H10.01c. SP (S. pombe), A (chromosome 1), cSH10
 (cosmid name), 01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions. Cosmid c23C11
 overlapped at the 3' end by c13F5.
 Location/Qualifiers
 1. 38391
 /organism="Schizosaccharomyces pombe"
 /strain="972h-"
 /db_xref="taxon:4896"
 /chromosome="I"
 /clone="cosmid c23C11"
 /map="IL"
 217..1799
 /gene="SPAC23C11.01"
 join(217..480,560..874,928..1113,1154..1266,1352..1799)
 /gene="SPAC23C11.01"
 /note="SPAC23C11.01; len:441aa; similarity: eg. to
 Y1090W, Y1090W, Y1090W, P40499, hypothetical 56.3 kd protein,
 (491aa), fasta scores, opt:268,E():1.4e-19, (24.78

identity: in 478 aa overlap)"
 /codon_start=1
 /protein_id="CAB11154.1"
 /db_xref="PID:c334129"
 /db_xref="PID:c334129"
 /db_xref="GI:2330765"
 /db_xref="GI:2330765"
 /db_xref="SWISS-PROT:O13909"
 /translation="MDQENSNVYDLSVSTVEDTSEFAMDSILEITIDDEAPIGKRP
 KANISTHFRNMLAMFNGEFLISGFVIVLVGIAFEIGRFGILITLALVEFFSA
 LKLFGLKRLATLHFEEDPLVFLILIALNASPIEONKYASFLASAMTALLHETP
 LFNLLGMAISLVVQALGHSRLVHNKSENMMFILLNASSAISMSLYLYVSSPS
 ISNPALMGESLAVTIVISIGVAGRAULSFASIMFLYATVYMAVCDFGNPMS
 SEKRPFDIPLPNILDSVHRLISTISATIDPRTIYNVLEFVMAAKYAPSVETFEER
 ISVAVATRLPAIONNIIFLEYSRTSKQDMISLSPCLIAVYTNLLQHLPTPSF
 TSPNQILCSAEIWRWVSALTLLVLAIELAVSKESDYGQALASHKRLD"
 481..486
 /gene="SPAC23C11.01"
 /note="gtaagt, splice donor sequence"
 540..559
 /gene="SPAC23C11.01"
 /note="ctacataaattaaatag, splice branch and acceptor"
 875..880
 /gene="SPAC23C11.01"
 /note="gtaagt, splice donor sequence"
 913..927
 /gene="SPAC23C11.01"
 /note="taacttaaacgcag, splice branch and acceptor"
 1114..1119
 /gene="SPAC23C11.01"
 /note="gtaagt, splice donor sequence"
 1141..1153
 /gene="SPAC23C11.01"
 /note="taactattag, splice branch and acceptor"
 1267..1272
 /gene="SPAC23C11.01"
 /note="gtaagt, splice donor sequence"
 1335..1351
 /gene="SPAC23C11.01"
 /note="taactatttaacag, splice branch and acceptor"
 complement(2103..2583)
 /note="partial primary transcript for ribosomal protein
 s23 homologue; EM_FUN:SPAB1288"
 complement(2103)
 /note="poly A attachment site"
 complement(2135..2566)
 /gene="rps23"
 complement(2135..2566)
 /gene="rps23"
 /note="SPAC23C11.02c; len:143aa"
 /codon_start=1
 /product="40s ribosomal protein s23"
 /protein_id="CAB11155.1"
 /db_xref="PID:c334130"
 /db_xref="PID:c334130"
 /db_xref="GI:2330766"
 /db_xref="GI:2330766"
 /db_xref="SWISS-PROT:P79057"
 /translation="MGKPAGLNARKLRNRRERADAHYKRLGTAYAKSPFGS
 SHAKGIVKIKIVGAKOPNSAIRKCVQVQLIKGKVTAFVPHDGLNFEDNDENVLL
 SCGRGRKAKGDIIGVRFVYKAGVGLSLFHEKKEKPPA"
 complement(2366..2389)
 /gene="rps23"
 /note="P500055 Ribosomal protein S12 signature"
 complement(2633..2640)
 /note="Homol D box"
 complement(2642..2651)
 /note="Homol E box"
 2923..4719
 /gene="SPAC23C11.03"
 2923..4719
 /gene="SPAC23C11.03"
 /note="SPAC23C11.03; len:598aa, similar eg. to Y1R002W,
 Y1R21EAST, P47083, hypothetical 67.0 kd protein, (593aa),
 fasta scores, opt:855,E():0, (33.08 identity in 610 aa

```

overlap"
/codon_start=1
/protein_id="CAB11156.1"
/db_xref="PID:e334131"
/db_xref="PID:g2330767"
/db_xref="GI:2330767"
/db_xref="SWISS-Prot:013910"
translation="MTELINSEPTLEFLPNKEDSKOLVERAKEIVSFNIKIPSTPIOT
ITVKGIDNMQIYNQIALAKYALDDLPVDVNEDEDEDSITSPODEDETSHOELSD
NSEMNRDADADELFGSEPELNEVDSEAEVDSGEDEKMEINNEGOOIOEKKDAFGL
NDGEPIDENKOTILEEVAEVEGLGDEDDIDLMDDPMENNESDEENROADS
IMTEDEPFKRTAGRELLKKKSKRRKSPDEPNDOAVNIENNESNENETATE
DRYKDLFASDEDDVSDOLSSEDKRLITQOIELDAEVAKSWTMEAEATSKG
RPSNLDLDLEETCAKYPVOTETITLEDLKNRISTFTFDVPRKRAVATEF
RSEBELLENNSORSIAEEIEEELKSNADITYSEADKKKEKEHEELKALFSEVR
TIDSLASWNVPPPESTVEIYVSNAPTLAMEVOPADTMAALPOEYKRSNRE
GETITRSGIAISTEMDHQOKARRRVRKRAEKROAKERKRNSTQVVOYRQJLKS
NVEVIGKGEKRVASNKYPHYVSSNQLKL"
complement(4842..6216)
/gene="SPAC23C11.04c"
complement(10114842..4988,5045..5622,5676..6216))
/gene="SPAC23C11.04c"
/note="SPAC23C11.04c; len:421aa, similar eg. to C. elegans
Q19683, F21D5.5, (250aa), fasta scores, opt:395,
E(1.1e-32, (41.7% identity in 223 aa:overlap)also
similar eg. to YMR156C, YMR31.YEAST_Q03796, hypothetical
27.4 kd protein, (238aa), fasta scores, opt:156,
E(1.4e-15, (32.3% identity in 206 aa:overlap), contains
P50017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/protein_id="CAB11157.1"
/db_xref="PID:e334132"
/db_xref="PID:g2330768"
/db_xref="GI:2330768"
/db_xref="SWISS-Prot:013911"
translation="MSSKRRKSPQESLTSYFEKSSKSKKYGSONKDDSSSTCLOO
KLEIOWSITDSLAIYAKYGLKRTKKEFLAFIDDLGTLIKRSGVSKDAWMTWHPRS
VFKALYODNYSIVFNONIGIPKPSAGHTFOKIRAITESDLPVLYAAILDKD
FRKPLTGMSNFKLVNRSIDSLPIKRYGDAAGRGDINSIDLKFAENIGIKFTPEQ
FELGSPVPNFSHPKNTLVNRSNPHYHKKSEHGVIVLGVFPSSGRTAESQ
IVTQGVIERNODILTKSKCIKAAIEALKKRSVIGWYSISTTYAISDNTLTIES
RKMWIDIAOEIEIPICHLQSEELARHNHVFRIYHNKOLPEIARNSKSRFQMP
TVEEGFVVEEPEKCLDYEDTWNYYE"
complement(4989..5002)
/gene="SPAC23C11.04c"
/note="c1aaccaattctag, splice branch and acceptor"
complement(5038..5044)
/gene="SPAC23C11.04c"
/note="gtaagt, splice donor sequence"
complement(5354..5377)
/gene="SPAC23C11.04c"
/note="PS00017 ATP/GMP-binding site motif A (P-loop)"
complement(5623..5643)
/gene="SPAC23C11.04c"
/note="c1aaccaattctag, splice branch and acceptor"
complement(5670..5675)
/gene="SPAC23C11.04c"
/note="gtaagt, splice donor sequence"
7157..8475
/note="EM-FUN:SPPPAG X54301 yeast ppa gene for inorganic
pyrophosphatase"
7198..7203
/note="putative TATA box"
7307..8176
/gene="ppal"

```

Note: remainder of annotations omitted.

Query Match 7.8%; Score 100; DB 27; Length 38391;
 Best Local Similarity 59.7%; Pred. No. 3.19e-54;
 Matches 308; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Db 7422 ATGCTGAGAACACATCTTGAAACATGGTGTGAGATCCCTGTTGAGACTCAGGCTAAGC 7481

```

OY 201 ATGACATATAGAGATGTGTTTACATGTAGTGTGAAGTACACCGCTGGTCTAATGCAAAA 260
Db 7482 TTGAATTTACCAAGAGAGTACTTGAACCTTATCAGCAGCATACTAAGAGGTTAAGC 7541
OY 261 TGGAGATTCTCAAAAGAGACCTTTAAACCTTATTAACCAAGATGTGAAAAAGAGAAAC 320
Db 7542 TTGCTTGGTGTGTAAGTCTTCTTCCACACAGGTTTACATTTGGAAGTACGCTTTTC 7601
OY 321 TTGGCTATTTCCGATTTGTTTCCGTTAAGGATATATCTGTAACATATGTCATCC 380
Db 7602 CTCAACCTATGAGATCTTAAATGTTTCAATCCGTAACCAAGGCAAGGCTGTATACG 7661
OY 381 CTGACAGTTGGGAAGACCAAGGCAACATATGATACATACCTGCTGTGTGTGATG 440
Db 7662 ATCCCTGATTTGTGTAGATCGGTGAAGCTGCTGTACACCGGCAAGTCAAGCAAG 7721
OY 441 ACCCAATTGATGTGTGTAATTTGGAACCAAGGATGTGCAAGAGGTGAATTAATGGCG 500
Db 7722 TTAAGTCTTGGGTGTATGTCCTTTCGACGAGGAGTGTGAGTGTGAAGTGTATG 7781
OY 501 TGAAGTTCTAGGCATATGCTGATGATTTGACGAGGGAACCGACTGGAAGTCAATG 560
Db 7782 TCATGACGTTAATGACCTTTAGCTCCTTAAGCTTAATGATATGAGAGTGTGAAGTC 7841
OY 561 CCATTAATGTGATGATCTGATGACGACCAATTAATGATATCAATGATGTCAACGCC 620
Db 7842 ATATCGGTGATTAATCGCTCTACTACAGAGTGTGTCGATTTACAAGATCCCGATG 7901
OY 621 TGAACCTGGCTACTTACAGCTACTGTGACTGTGTTAAGGATTAAGGTTGCTGATG 680
Db 7902 GCAACCTGAAATTCCTTTCCTTTCAGTGTGAAT 7937
OY 681 GAAACCAAGAAATGAGTTCCGTTTATGACGAT 716

```

RESULT 12.

LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM
AF085600	2826 bp DNA	AF085600	93746879	AF085600.1	GI:3746879	fruit fly.	Drosophila melanogaster
AF085600	Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38)	AF085600	93746879	AF085600.1	GI:3746879	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	Drosophila
1	(bases 1 to 2826)						Gula,D.A., Sandatizopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.
2	(bases 1 to 2826)						Gula,D.A., Sandatizopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.
3	Inorganic pyrophosphatase is a component of the Drosophila nucleosome remodeling factor complex						
4	Submitted 37 Room 5E-20, Bethesda, MD 20892-4255, USA						
5	Building 37 Room 5E-20, Bethesda, MD 20892-4255, USA						
6	Location/Qualifiers						
7	1..2826						
8	/organism="Drosophila melanogaster"						
9	/db_xref="taxon:7227"						
10	/chromosome="2"						
11	/map="60c"						
12	join(554..704,1148..1267,1333..1728,1784..1935,1994..2102,2422..2585)						
13	/gene="Nurf-38"						
14	/product="NURF-38"						
15	554..2585						
16	/gene="Nurf-38"						
17	join(642..704,1148..1267,1333..1728,1784..1935,1994..2102,						

Db 1064 GATYY 1123

FEATURES

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
Cosmid c3A12 overlapped by c20G8 at the 5' end.
Location/Qualifiers

```

source
1. 37013
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/clone="cosmid c3A12"
/map="IL"
/complement(1. 1114)
/gene="SPAC3A12.01"
1. 110
/gene="SPAC3A12.01"
/note="nominal overlap with c20G8"
/complement(1. 1114)
/partial
/gene="SPAC3A12.01"
/note="SPAC3A12.01c, similarity: to YPL120W, Q02948,
LPH7P, (557aa), fasta scores, opt:391, E():4.9e-16, (25.68
identity in 344 aa overlap)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB08758.1"
/db_xref="PID:e351362"
/db_xref="PIR:G2462674"
/db_xref="GI:2462674"
/db_xref="SWISS-PROT:P87117"
/translation="MOYLCOCHSLINKKDYDDDLKIKLPLKSPFYQASLTENME
SGESDDDNSSSEDPAPQRLQYKKTISEGYNNDNPPPLRTPTLDSFVILPAAD
GYEEKNSPEPVNLFSSKITEYRIPDLSSKRVVHPFLCYEACALLITEMSKTLA
LKEKKMYFNDNLSTQVTEENPAIDSDIDELMKOINKKEKIEISDETLOK
LRLDEKEVVAEEDFVNNLQFOFKLSTLRQYDCALFEHNSRKLEKLOKKN
VFSDIYISHSEPNESGIATINGLRGRGPSQKVMVAEINAMGMVILLVLTKE
LDFHSSYOLKPFQSGSFIIRDRDPNGNQYKPKLD"
1796. 1868
/note="tRNA-Cys, anticodon gca, length = 73"
2563. 3589
/gene="SPAC3A12.02"
/join(2563. 2718, 2827. 3199, 3258. 3589)
/gene="SPAC3A12.02"
/note="SPAC3A12.02, len:286aa, similar eg. to IPYR_SCHPO,
P19117, inorganic pyrophosphatase, (286 aa), fasta scores,
opt: 930, E():0, (48.78 identity in 277 aa overlap), also
similar eg. to YBR011C, IPYR_YEAS, P00817, inorganic
pyrophosphatase, (286 aa), fasta scores, opt: 904, E():0,
(54.4 identity in 250 aa overlap)"
/codon_start=1
/product="inorganic pyrophosphatase"
/protein_id="CAB08747.1"
/db_xref="PID:e316097"
/db_xref="PIR:G2104418"
/db_xref="GI:2104418"
/db_xref="SWISS-PROT:P87118"
/translation="MASLAKNLIQFRSKIKINTPPRYVCYKNNKPISEFHDVPLT
SDKRTVMWTEIPWTOAKCEISLTSFHPKIDLAKGKLRVANSFPIGFHWNGA
LPQWEPPNVIDSKTKMGGDPLDVCIEIGSISGILQIKQVAKGLIDGENDTGA
KILAIIDINDPRAKILNDISDVQNLMPRLCTCRWFPIYIIPCKPKNRFPGNYLP
KSAOLDIIOACHQHMWKSBRKOYIKNFHNSVNMVLLINKINSIKKEVQNSNYS
FPHYHILPIL"
2713. 3267
/gene="SPAC3A12.02"
/note="pfam match to entry PF007719 Pyrophosphatase,
Inorganic pyrophosphatase"
2719. 2724
/gene="SPAC3A12.02"
/note="gstatgt, splice donor sequence"
2806. 2826
/gene="SPAC3A12.02"
/note="ctaacatttttcaactag, splice branch and acceptor
sequence"
3034. 3054
/gene="SPAC3A12.02"
/note="PS00387 Inorganic pyrophosphatase signature"
3200. 3205
/gene="SPAC3A12.02"

misc_feature
3245. 3257
/note="gtagct, splice donor sequence"
/gene="SPAC3A12.02"
/note="ctaacattcttag, splice branch and acceptor sequence"
complement(3887. 4816)
/gene="SPAC3A12.03c"
complement(3887. 4816)
/gene="SPAC3A12.03c"
len:309aa, similar eg. to g_gallus
090972, RING ZINC FINGER PROTEIN, (381 aa), fasta scores,
opt:214, E():1.4e-07, (37.3 identity in 67 aa overlap)
also similar to GOL1_DROME, Q06003, gollath protein (91
protein), (284 aa), fasta scores, opt:203, E():6e-07,
(36.5 identity in 74 aa overlap)"
/codon_start=1
/product="zinc finger, C3HC4 type"
/protein_id="CAB08748.1"
/db_xref="PID:e315874"
/db_xref="PIR:G2104419"
/db_xref="GI:2104419"
/db_xref="SPTRMBL:P87119"
/translation="MPIPKRPMHVEEVEVSSQNTTEILLFALYIISVIRPFEEFL
RCQCVYFHTLENOEGDDEPRLIOHNVNSTGSLSPVRLNVLGYDIPSRRSV
VSKKALSCISLEIPIYIKMLKKRGHAKGSESTPVSNSQSVYVOGGEPPSYITTD
VRBPNGSTSFVEMSSLSNITNTDASDSDSDSCLEDEDFCTIADYAFDIL
RVLPCEBVFHTQCIDTWITTKASCPLCNDSDYKYLPLQMAASVTHENAMSIPLSP
GDSRTHAEITRSLLSMSVSRNSRMPYVYSTL"
complement(4079. 4204)
/gene="SPAC3A12.03c"
/note="pfam match to entry PF00097 zf-C3HC4, zinc finger,
C3HC4 type"
3344 type"
complement(5344. 6105)
/gene="SPAC3A12.04c"
complement(join(5344. 5808, 5878. 5976, 6019. 6105))
/gene="SPAC3A12.04c"
/note="SPAC3A12.04c, len:216aa, similar eg. to YHR062C,
YHM2_YEAS, P38786, hypothetical 32.2 kd protein, (293aa),
fasta scores, opt:420, E():4e-22, (34.9 identity in 218
aa overlap)"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAB08749.1"
/db_xref="PID:e316098"
/db_xref="PIR:G2104420"
/db_xref="GI:2104420"
/db_xref="SWISS-PROT:P87120"
/translation="MINFYLOKVGTAIALANQYDGKLONVKINPIKVEKPEOKK
IYSRITLITLESMPQNVLSNVTKPEPILAIRPGDILLQOTSDLEFDLISDFQRL
PFYLRKTFMGIAVSRDGIETISYSGISLRVSNRNITNATISVYRATRRGITVSET
KPILECRAGFDVNLNATFMDLKODOKARKVSGSCRSVLLHAETRRDTYSILNGCH"
complement(5809. 5823)
/gene="SPAC3A12.04c"
/note="ctaacattgccttag, splice branch and acceptor
sequence"
complement(5872. 5877)
/gene="SPAC3A12.04c"
/note="gtagct, splice donor sequence"
complement(5977. 5985)
/gene="SPAC3A12.04c"
/note="ctaacatttag, splice branch and acceptor sequence"
complement(6013. 6018)
/gene="SPAC3A12.04c"
/note="gtagtg, splice donor sequence"
complement(6860. 10301)
/gene="SPAC3A12.05c"
complement(join(6860. 7795, 7836. 10215, 10291. 10301))
/gene="SPAC3A12.05c"
/note="SPAC3A12.05c, putative transcription initiation
factor, len:1108aa, similar eg. to YCR42C, T2D2_YEAS,
P23253, transcription initiation factor tsm1, (1407aa),
fasta scores, opt:1598, E():0, (31.0 identity in 1226 aa
overlap)"
/codon_start=1

```


[illegible]

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MSPorch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Dec 27 12:53:55 1999; MasPar time 277.57 Seconds
Tabular output not generated. 984.229 Million cell updates/sec

```

Title:	US-09-415-540-2
Description:	(1-1275) from US09415440.seq
Perfect score:	1274
N.A. Sequence:	1 CAAGAGCTTNGGGGCTCTCT.....GAATTAATTAATTTTGGTG 1275

Scoring table: TABLE default

Nmatch STD : Dbase 0; Query 0

Post-processing: Minimum Match 0%

Database:

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.230; Variance 5.556; scale 1.661

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	pred. No.
1	1274	100.0	1275	54	V63542	cDNA encoding a human	0.00e+00
2	1265	99.3	1351	60	X00685	Human secreted protein	0.00e+00
3	1258	98.7	1350	60	X00716	Human secreted protein	0.00e+00
4	330	30.6	416	19	T21836	Human gene signature	2.18e-24
5	44	3.5	91	9	O51746	Oligonucleotide probe	2.95e-11
6	42	3.3	204	1	N81164	Basse substituted E.co	4.25e-10
7	40	3.1	91	9	O51746	Oligonucleotide probe	5.93e-09
8	40	3.1	204	1	N81164	Basse substituted E.co	5.93e-09
9	36	2.8	114	12	O70465	Genetic DNA sequence	1.04e-06

C	10	36	2.8	114.12	070468	Generic DNA sequence	1.04e-06
C	11	36	2.8	169.32	T76307	Human RANTRF antisens	1.04e-06
C	12	34	2.7	91.46	V44650	Mammalian DNA replica	1.29e-05
C	13	35	2.7	114.12	070469	Generic DNA sequence	3.68e-06
C	14	34	2.7	114.12	070467	Generic DNA sequence	1.29e-05
C	15	34	2.7	114.12	070467	Generic DNA sequence	1.29e-05
C	16	33	2.6	114.12	070468	Generic DNA sequence	4.47e-05
C	17	32	2.5	91.46	V44650	Mammalian DNA replica	1.53e-04
C	18	32	2.5	114.12	070470	Generic DNA sequence	1.53e-04
C	19	32	2.5	114.12	070469	Generic DNA sequence	1.53e-04
C	20	32	2.5	114.12	070465	Generic DNA sequence	1.53e-04
C	21	32	2.5	114.12	070466	Generic DNA sequence	1.53e-04
C	22	32	2.5	190.32	T76452	Chymase antisense oli	1.53e-04
C	23	31	2.4	114.12	070466	Generic DNA sequence	1.72e-03
C	24	30	2.4	114.12	070470	Generic DNA sequence	1.72e-03
C	25	30	2.4	114.12	070472	Generic DNA sequence	1.72e-03
C	26	31	2.4	117.32	T76363	Human interleukin 8 a	5.17e-04
C	27	30	2.4	178.32	T76405	Human endothelin-1 an	1.72e-03
C	28	30	2.4	657.7	043519	Degenerate FMN reduct	1.72e-03
C	29	29	2.3	70.32	T76184	Human IL4 antisense o	5.66e-03
C	30	29	2.3	178.32	T76605	Human endothelin-1 an	5.66e-03
C	31	28	2.2	89.32	T76219	Human IL6 antisense o	1.83e-02
C	32	28	2.2	128.32	T76233	Human IL6 antisense o	1.83e-02
C	33	28	2.2	160.47	V48098	Oligonucleotide Lp160	1.83e-02
C	34	28	2.2	250.32	T76438	Substance P antisense	1.83e-02
C	35	27	2.1	88.32	T76170	Human IL3 receptor an	5.80e-02
C	36	27	2.1	114.12	070472	Generic DNA sequence	5.80e-02
C	37	27	2.1	114.12	070471	Generic DNA sequence	5.80e-02
C	38	27	2.1	130.47	V48104	Randomised Pool oligo	5.80e-02
C	39	27	2.1	160.47	V48098	Oligonucleotide Lp160	5.80e-02
C	40	26	2.0	114.12	070473	Generic DNA sequence	1.81e-01
C	41	26	2.0	140.47	V48104	Randomised Pool oligo	1.81e-01
C	42	26	2.0	136.21	T13624	MW2b library generat	1.81e-01
C	43	25	2.0	264.32	T76445	Substance P receptor	5.53e-01
C	44	26	2.0	501.3	N50033	Sequence encoding new	1.81e-01
C	45	26	2.0	2539.51	V59609	Human secreted protei	1.81e-01

ALIGNMENTS

RESULT	1
ID	V63542 standard; cDNA; 1275 BP.

DT 03-FEB-1999 (first entry)

KW Human; Inorganic pyrophosphatase; HPYP; Incyte clone 768320;

OS Homo sapiens.

FT	CDS	89.958
----	-----	--------

FT
/product= HPYF

PD 01-DEC-1998

PR 31-OCT-1996; US-74
PA (INCY-) INCYTE PH

PL HAWKINS PK, HILLMAN
DB WPT: 99-044570/0A

PT DNA encoding human

PT recombinant protein tissue regeneration

CLAIM 3, FIG. 3A-D

The present sequence

CC designated HPIF. 1
CC in Incyte clone 70

and tissue regeneration

[illegible]

Best Local Similarity

Query Match	100.0%;	Score 1274;	DB 54;	Length 1275;
Best Local Similarity	100.0%;	Pred. NO.	0.00e+00;	

Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 1 caaaggttnggggtctctctctctctctcaagtcggcgccgctgagggcttggtctcttg 60
Oy 1 CAAGAGGTTNGGGGCTCTCTCTCTCTCTCAAGTCCGCCGCCGCTGGCTGTG 60
Db 61 gcaagggcgcgccgaggaactccggagactatgagcggttccaccaggagagcgccgc 120
Oy 61 GCAGGGCGCGCGGAGAGACTCCGGCACTATGAGCGGCTTACACCGAGAGAGCGCGCC 120
Db 121 gcccttcccttgagatccgagcttccctcaaaaaatgaaaggacatatatctcc 180
Oy 121 GCCCTTCCCTCGAGATCCGAGTCTCTCAAAAATGGAAGGCAATATATCTCC 180
Db 181 attcatatltccaattttagcagataagatgtgttcaatggttagtgaagttacc 240
Oy 181 ATTCTATATATTCCAATTTATGCAATTAAGATGATGTGTTTCACATGTAGTGAAGTACC 240
Db 241 aagctggtctaatgcaaaaatgagatttgctacaaaggaccccttaaacctattaaaca 300
Oy 241 AAGCTGGTCTAATGCAAAAATGGAATTTGCTACAAAGACCTTTAAACCTATTAAACA 300
Db 301 agatgtgaaaaaaggaaacttcgctatgttgcaattgttcccgatlaaaggatata 360
Oy 301 AGATGTGAAAAAAGGAAACTTCGCTATGTTCGAATTTGTTCCGTAATAAGATATAT 360
Db 361 ctggaactatgtgtccatcccgagacttgggaagaccgaaggacaaatgataaacatcc 420
Oy 361 CTGGAACATATGTGTCCATCCCTGACACTTGGGAGACCAGGCGCAATGATAAACATAC 420
Db 421 tggcgtgtgtgtgacaaatgacccaattgattgtgtgtaaatggaagaagatgtg 480
Oy 421 TGGCGTGTGTGTGACAAATGACCCCAATGATGTGTGTAATGGAAGCAAGATATGTC 480
Db 481 aagaggtgaataaattggcggtgaaagttctagcgaattgtgctagatgagaaaggga 540
Oy 481 AAGAGGTGAATAAATTGGCGGTGAAGTTCTAGGCATATTGGCTATGATTGAGAAAGGGA 540
Db 541 aacggactgggaagatcatctgccatlaaagtgtgaatgccctgaatgagaaaggga 600
Oy 541 AACGGACTGGGAAGATCATCTGCCATTAATGTGATATGATCTGATGACCAATTAATGA 600
Db 601 tatcaatgatgtcaaaagcctggaaccctggacttagaagctatgtggactgtttag 660
Oy 601 TATCAATGATGTCAAAAGCCTGGAACCTGGCTACTTAAAGCTACTGTGGACTGATTTAG 660
Db 661 aaggtataagttccctgtgtgaaacccgaaatagatttgcgtttaaagaaattaa 720
Oy 661 AAGGTATAGGTTCTCTGATGGAACCAAGAAATGAGTTGGGTTAATGCAATTTAA 720
Db 721 agataaggacttggcattgatattatataaagccctatgaccttggaaagcattagt 780
Oy 721 AGATAGGACTTTGGCATTGATATTATTAAGCACTATGACCAATTTGAAAGCAATTTAGT 780
Db 781 gactaagaagaacgaatgaaagaagaatcagttgcaatgaatacaaatgtgttgaagccc 840
Oy 781 GACTAAGAAACGAATGAAAAAGCAATAGTTGCATGATATCAAACTTTGTGGAAGCCC 840
Db 841 ctccaagtgtatccctgtgtctgcagaagcatttggatgtgattacacaccccttga 900
Oy 841 CTTCAAGTGTATCCTGTGATGCTGCAGAGCAATGTGATGCTTTACACACCCTGTGA 900
Db 901 atctgctgtacagtaaccaagaacggtgataaagtgttccatcaaccagaanaataatg 960
Oy 901 ATCTGCTGTACAGTACCAACAGACGATGATAGGTTTCATCACCAGAAAACCTAATG 960
Db 961 agatttcttggaataaagatgtatgtctacacatcggttccatctgtgattatagaag 1020
Oy 961 AGATTCTCTGGAATAAAGATGTGATGTCTACACATCGTTCACTGTGATATTATTAAG 1020
Db 1021 taaaagtagtagcttctcaaaagctttaaattttagaactcatctaaactaaagtaaatc 1080
Oy 1021 TAAAGTAGTAGCTTCTCAAGCTTTAAATTTGTAGAACCTCATCTAACTAAGTAAATTC 1080

```

Db 1081 tgcgtgactaatccaatalactcagaatgttatccatctaaagcattttcatatcca 1140
Oy 1081 TGCTGTGACTAATCCAATAATCTGAGATGATATCATCTAATGATTTTCATATCCCA 1140
Db 1141 actaagaatacttttaagacatgctttaaataatcaagcagttgtcatttggaaatcact 1200
Oy 1141 ACTAAGATACCTTTTAAAGCAATGCTTAATATCAAAAGCATTTGTCATTGGAAAGTCACTT 1200
Db 1201 gtgaataatgtgcaaggaggacacatatgtgatgtatatttaccatgttgaagaa 1260
Oy 1201 GTGAATAATGTGCAAGGAGGACACATTTGATGTATATGTTACCATATGTTAGAAAT 1260
Db 1261 aaataatttgcgtg 1275
Oy 1261 AAAATATTGTCGTG 1275

RESULT 2
ID X00685 standard; DNA; 1351 BP.
AC X00685;
DT 25-MAR-1999 (first entry)
DE Human secreted protein gene 75 clone H0SE145.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN W09842738-A1.
PD 01-OCT-1998.
PE 19-MAR-1998; U05311.
PR 30-MAY-1997; US-050937.
PR 21-MAR-1997; US-041276.
PR 21-MAR-1997; US-041277.
PR 21-MAR-1997; US-041281.
PR 21-MAR-1997; US-042344.
PR 30-MAY-1997; US-048069.
PR 30-MAY-1997; US-048094.
PR 30-MAY-1997; US-048095.
PR 30-MAY-1997; US-048096.
PR 30-MAY-1997; US-048099.
PR 30-MAY-1997; US-048131.
PR 30-MAY-1997; US-048135.
PR 30-MAY-1997; US-048154.
PR 30-MAY-1997; US-048160.
PR 30-MAY-1997; US-048186.
PR 30-MAY-1997; US-048187.
PR 30-MAY-1997; US-048188.
PR 30-MAY-1997; US-048350.
PR 30-MAY-1997; US-048351.
PR 30-MAY-1997; US-048352.
PR 30-MAY-1997; US-048353.
PR 05-AUG-1997; US-054804.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Ferlie AM, Florence KA,
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,
PI Rosen CA, Ruden SM, Shi Y, Young P;
DR WPI: 99-070066/06.
DR P-PSDB: W67881.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PS disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 233-234; 385pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.

[illegible]

RESULT	3	
ID	X00716	standard; DNA; 1350 BP.
AC	X00716:	
DT	25-MAR-1999	(first entry)
DE	Human secreted protein gene 75 clone H05E145.	
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;	
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;	
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;	
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;	
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
OS	Homo sapiens.	
PN	W09842738-A1.	
PD	01-OCT-1998.	
PR	19-MAR-1998:	U05311.
PR	30-MAY-1997:	US-050937.
PR	21-MAR-1997:	US-041276.
PR	21-MAR-1997:	US-041277.
PR	21-MAR-1997:	US-041281.
PR	21-MAR-1997:	US-042344.
PR	30-MAY-1997:	US-048069.
PR	30-MAY-1997:	US-048094.
PR	30-MAY-1997:	US-048095.
PR	30-MAY-1997:	US-048096.
PR	30-MAY-1997:	US-048099.
PR	30-MAY-1997:	US-048131.
PR	30-MAY-1997:	US-048135.
PR	30-MAY-1997:	US-048154.
PR	30-MAY-1997:	US-048160.
PR	30-MAY-1997:	US-048185.
PR	30-MAY-1997:	US-048187.
PR	30-MAY-1997:	US-048188.
PR	30-MAY-1997:	US-048350.
PR	30-MAY-1997:	US-048351.
PR	30-MAY-1997:	US-048352.
PR	30-MAY-1997:	US-048355.
PR	05-AUG-1997:	US-054804.
PA	(H05A) - HUMAN GENOME SCI INC.	
PI	Breene LA, Huan R, Ebner R, Ferrie AM, Florence KA,	
PI	Greene JM, Du J, Lallier DW, Moore PA, Ni J, Olsen HS,	

Pt	Rosen CA,Ruben SM, Shi Y, Young P:
Df	WPI: 99-070066/06.
DR	P-PSDB; W67912.
PT	New isolated human genes and the secreted polypeptides they encode
CC	diseul for diagnosis and treatment of e.g. cancers, neurological
PT	disorder, immune diseases, inflammation or blood disorders
PS	Claim 1; Page 264-265; 385pp: English.
CC	This sequence represents a nucleic acid molecule which encodes a secreted
CC	human protein. The gene number, and the clone it is derived from, are
CC	detailed in the descriptor line. The gene can be used to generate fusion
CC	proteins by linking to the gene to a human immunoglobulin Fc portion
CC	(e.g. X00602) for increasing the stability of the fused protein as
CC	compared to the human protein only.
CC	The invention relates to 87 novel genes and their fragments (nucleic acid
CC	sequences: X00611-X00724, amino acid sequences W67807-W68004) which
CC	are useful for preventing, treating or ameliorating medical conditions
CC	e.g. by protein or gene therapy. Also, pathological conditions can be
CC	diagnosed by determining the amount of the new polypeptides in a sample
CC	or by determining the presence of mutations in the new polynucleotides.
CC	Specific uses are described for each of the 87 polynucleotides, based on
CC	which tissues they are most highly expressed in (see X00611 for described
CC	uses).
SQ	Sequence 1350 BP; 414 A; 250 C; 319 G; 363 T;
Query Match	98.7%; Score 1258; DB 60; Length 1350;
Best Local Similarity	99.9%; Pred. No. 0.00e+00;
Matches 1264; Conservative	0; Mismatches 0; Indels 1; Gaps 1;
Df	20 ggggctctctctttgttcagtcgcgcgcgcgtcgcgctgtgtgtctctgttgcaagcgcg 79
Oy	11 GGGGCTCTTCCTTGTCAGTGC GGCGCGGTGGGGCTGTGGCTCTGTGGAGCGCGG .70
Df	80 cggcaggactccggcactatlgagcggcttcacgccagaccgagagcgcgcgcc -ttctcc 138
Oy	71 CGCGAGACATCCGGCACTATGACGGCCTTAGACCAGCAGAGACGGCGGCCCTTTCC 130
Df	139 ctggagracccaagtcttcctccaaatatgagaagacaataatattctccattcatgat 198
Oy	131 CTGGAGTACCGAGCTTCCTCCAAATAATGAGAAGACAATAATATTCCTCATTCATGAT 190
Df	199 attccaattatgcagataagagatgtgttcacatgttatgtagtcccacgctgacct 258
Oy	191 ATTCCATTATTATGAGATAGATAGATGTGTTTCACATGTGATGTAAGTACACGCTGGCT 250
Df	259 aaatgcaaataatgagatltgtctacaagaagcccttaaacctataaacaagatgtgaa 318
Oy	251 AAATGCAAAAATGAGATGCTGTACAAAGAGACCCCTTAAACCTATTAAACAGATGTGAAA 310
Df	319 aaaagaanaacttgcatafttgtgaatttgttccgtataaagatatatctcggaacct 378
Oy	311 AAAAGAAAACCTTGCTATGTGTGGAAATTTGTCCTCCGTATAAAGATATATCTGGAACTAT 370
Df	379 gtgtccatccctcagacttggagaagccccagagcacaaatgataaacatactgctgtgt 438
Oy	371 GTGTCCATCCCTCAGACTTGGAAGACCCAGGGCAAAATGATAAACATATCTGCTGTGT 430
Df	439 ggtgacaatgacccaattgtatgtgttgaaatttggaagaagatgtgtgcaagagtgtaa 498
Oy	431 GGTGTACATATACCCAATGTATGTGTGAAATTTGGAACCAAGTATGTGCAAGAGTGA 490
Df	499 ataattgsgtgaagaagtcttagagcatattgttgcatatgtgaagaagggaacagactg 558
Oy	491 ATAATTGTGGCTGAAGAGTTCTTAGGCATATTTGGCTATGATGACGAAGGGGAAAAACCCACAGCG 550
Df	559 aaagtcactgcataatgtgagatgtacctgtatgacgcaccaatcataatgataataaat 618
Oy	551 AAAGTCATTGCCATTATGTGTGATGATCTGTATGACGCCAATTAATAATGATATCAATAT 610
Df	619 gtcaaaagcgtgaacacctgtgctaatagaagcactactgtgactgtttagaaggtataag 678
Oy	611 GTCAAAAGCGGTGAACACTGTGCTACTTAAAGAGCTACTGTGACTGGTTTGAAGGTATAG 670
Df	679 gtctccagtgaagaacccagaataatgagittgcgtttaatgcagaatattaagaataagac 738

QY	671	GTTCCTATGGAAGAACAGAAATGAGTTGGCTTTAATGCGAATTTAAAGATTAAGAC	730
Db	739	tttgccatgatatttaaaagcactcagccattggaaagcattagtgactaagaa	798
QY	731	TTTSCCATGATATTTAATAAACCACTCATGACATGTGAAGCATTTAGTGACTAGAAA	790
Db	799	acgaatggaaaaggaatcagttgatcatgaaatacaacttgcctgtagagccctcctaagt	858
QY	791	ACGAAATGAAAAGAAATCACTGTTCCATGAAATACAACTTGTCTGAGAGGCCCTTCAAGTGT	850
Db	859	gattcctatgctgcacagacattgtagtgctttacccaccctctggaatctgctgc	918
QY	851	GATCCTATCTGCGACAGGCACTTGGAGAGCTTTACACCAACCTGTGAATCTGCTGC	910
Db	919	acagtaaccaacagacgttggataagttgttccatcacacagaaaacataatgattcct	978
QY	911	ACAATACCAACAGACGCGTAGTAGTGTTCCATCAACCAAAATAATGAGATTCTCT	970
Db	979	ggaatacaagctgatattgctacatcgtgttcatctgtagtattagaagtaaaagtagt	1038
QY	971	GGAATACAACTCATATTGCTATCATCGTTCATCTCGATGATGAAAGTAAGTGTGT	1030
Db	1039	agctttcaagctttaaatctttagaactcctcaactaaagtaaatcttgccttgc	1098
QY	1031	AGCTTTCAAGCTTTAATTTTGTAGAACTCATCTTACTAAAGTAAATTTGCTGTGCT	1090
Db	1099	aatcaatactaccagaatgttcatcatcatcaaaagcattttcaatctcaactaagataa	1158
QY	1091	AATCAGATATACACAGATGTTATTCATCTAAAGCATTTTTCATATCTCACTAAGATA	1150
Db	1159	cttttagcacatgcttaaatatcaatacaagcagttgtcaatttggaaagcacttgyaatagat	1218
QY	1151	CTTTTACACATGCTTAAATATCAAAAGCAGTTGTCATTTGGAAGCACTTGTCGATATAT	1210
Db	1219	gtggaaggaggagacacattgtagtattgttaccatattgttagaataataatattt	1278
QY	1211	GTGAAAGGGAGACAAATTGATGTATATGTTACCATATGTTAGGAATAATAATATT	1270
Db	1279	tgctg 1283	
QY	1271	TGCTG 1275	
RESULT	4		
ID	T21836	standard: cDNA to mRNA; 416 BP.	
AC	T21836;		
DT	01-AUG-1996	(first entry)	
DE	Human gene signature HDWG503377.		
KW	Gene signature; messenger RNA; mRNA; relative abundance; frequency;		
KW	human; cloning; mapping; non-biased library; diagnosis; detection;		
KW	cell typing; abnormal cell function; ss.		
OS	Homo sapiens.		
PN	MO9514772-A1.		
DD	01-JUN-1995.		
EF	11-NOV-1994; J01916.		
PR	12-NOV-1993; JP-355504.		
PA	(MATS/) MATSUBARA K.		
PA	(OKUB/) OKUBO K.		
PI	Matsubara K. Okubo K.		
DR	WPI; 95-206931/27.		
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.		
PT	for diagnosis of abnormal cell function, by preparing cDNA that		
PT	reflects relative abundance of corresp. mRNA in specific human		
PT	tissues		
PS	Claim 1, Page 977; 2245pp; Japanese.		
CC	A single-stranded DNA (or its complementary strand or the corresp.		
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences		
CC	given in 119001-T26837 and which is able to hybridise to part of		
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)		
CC	sequences were obtained from 3'-directed cDNA libraries prepared		
CC	from various human tissues; synthesis of cDNA was initiated from the		
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-		

CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 416 BP; 128 A; 77 C; 77 G; 128 T;

Query Match 30.6%; Score 390; DB 19; Length 416;
 Best Local Similarity 97.6%; Pred. No. 2.18e-247;
 Matches: 405; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

Db 1 gatccgagtcgcccagagccattggtgagcttaccaccaccctgtaattcctc 60
 |||
 QY 851 GATCCGAGTCGCCCAGAGCCATTGAGTGCCTTACCCACCCTGTGATTCGCTGC 910
 Db 61 acagtcacacagagctgataagtggtccatccaccagaaacatagattctctc 120
 |||
 QY 911 ACAGTACCAACAGACCTGCATAGTGTTCCATCACCAGAAACTAATGAGATTCTCT 970
 Db 121 ggaatacagctgataattgctacatcggttccatctgtagtattagaagtaagtagt 180
 |||
 QY 971 GGAATCAAGCTGATTTGCTACATCGTTCATCTGGATGTTTGAAGTAAAGTAGT 1030
 Db 181 agctttcaagctttaattgttagaactcatcctaagaagtaattcgtcgtgact 240
 |||
 QY 1031 AGCTTTCAAGCTTTAAATTTGTAGAACCATCTAATCAATTAATTCCTGTGACT 1090
 Db 241 aatccatatctacagaaatgttatccatcctaagaacattttctatctcacaagaata 300
 |||
 QY 1091 AATCCATATFACCAATAGTTTATCCATCTAAGCAATTTTCTATCTCACTAAGATAA 1150
 Db 301 ctttagcacaatgcttaaatatcaataaagcagtggtcatttggaaagcactgtgaatagat 360
 |||
 QY 1151 CTTTAGCACAATGCTTAAATATCAAAAGCAGTTGTCAATTTGGAAGTCACTGTGAATAGAT 1210
 Db 361 ggcgaagccgcagacatatgtgtatattgtttccentatttggggataaa 415
 |||
 QY 1211 GTGCAAGGG-GAGCACATATGTGATGTATATGTT-ACCATATGTTAGGAATAAA 1263

RESULT 5
 ID 051746 standard; cDNA: 91 BP.
 AC 051746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 DR WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in
 PT samples
 PS Clam 3; Page 14; 23pp; English.
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
 CC (051735). It hybridized to all spp. of mycobacteria tested, but
 CC cross reacted to a few non-mycobacterial spp. The probe may
 CC be useful as an initial screen for mycobacterial infection.
 CC See also 051735-45 and 051747-59
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.58; Score 44; DB 9; Length 91;
 Best Local Similarity 5.6%; Pred. No. 2.95e-11;
 Matches 3; Conservative 46; Mismatches 5; Indels 0; Gaps 0;

Db 7 ggcgssvhsyvvvvhvshhsyvvvvhvshhsyvvvvhvshhsyvvvvhvshhsyvv 60
 |||
 QY 371 GGTCCATCCCTCAGACTTGGAAGACCCAGGACCAATGATTAACATCTGCG 424

RESULT 6
 ID N81164 standard; DNA: 204 BP.
 AC N81164;
 DT 08-NOV-1990 (first entry)
 DE Base substituted E.coli beta-galactosidase alpha-fragment.
 KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
 OS Escherichia coli.
 FH Key
 FT misc_feature
 FT 19.69
 FT /tag- a
 FT /function= multiple cloning site
 FT 187..204
 FT primer_bind
 FT /*tag- b

PN EP-285123-A.
 PD 05-MAY-1988.
 PE 30-MAR-1988; 105163.
 PR 03-APR-1987; US-034819.
 PA (SUSO) SUOMEN SOKERI OY.
 PI Lehtovaara P, Knowles J, Kotivola A, Bamford J, Reinkainen T;
 DR WPI: 88-279927/40.
 PT Introducing random point mutations into nucleic acids -
 PT by prep of single stranded template, annealing a primer, elongation,
 PT misincorporation, completion of molecules and screening.
 PS Disclosure; P; English.
 CC Random point mutations were introduced into the alpha fragment of
 CC E.coli beta-galactosidase. The wild type sequence was obtained as a
 CC single stranded template and an oligonucleotide was hybridised to
 CC it to generate a popn of DNA molecules which terminate at all
 CC possible nucleotide positions within a specified region. The
 CC variable 3' ends generated in this way are used as primers for
 CC reverse transcriptase. Nucleotides are misincorporated by the
 CC transcriptase and the molecules are completed to forms that can be
 CC amplified and then expressed in a suitable host-vector system.
 CC The sequence covers all 176 diff base substitutions, most of which
 CC occurred singularly in any given mutant.
 CC See also P80575.
 SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.38; Score 42; DB 1; Length 204;
 Best Local Similarity 11.0%; Pred. No. 4.25e-10;
 Matches 11; Conservative 53; Mismatches 36; Indels 0; Gaps 0;

Db 92 hbyrmbvnydyrdsaaaycyrrsvkydcyachdhhyvbbvnyvnhnc 151
 |||
 QY 282 CTTTAAACCTATTAAACAAGATGTGAAAAAGAAAACCTTCGTATGTGCAATTGT 341
 Db 152 ncccbnhvchvbnhbnhwayvzhdarrdvhcvcchc 191
 |||
 QY 342 TCCCGTAAAGATATATCTGGAAGTATGTCCTCC 381

RESULT 7
 ID 051746 standard; cDNA: 91 BP.
 AC 051746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 DR WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in

[illegible]

PS Example 5: Page 35: 71pp: English.

CC A method for treating airway disease in a subject has been produced,

CC which involves the topical administration of an essentially adenosine

CC free antisense oligonucleotide (ON) to the airway epithelium of the

CC subject. The present sequence is an antisense oligonucleotide specific

CC for the human RANP5, targeted at the initiation codon. The method can

CC be used to treat airway diseases such as cystic fibrosis, asthma,

CC chronic obstructive pulmonary disease, bronchitis and other airway

CC diseases characterised by an inflammatory response. By eliminating

CC adenosine from the antisense ON, its liberation upon antisense

CC degradation is prevented, thereby preventing adenosine-induced

CC bronchoconstriction in patients with hyper-reactive airways.

CC Sequence 162 BP; 0 A; 29 C; 74 G; 22 T;

SO

Query Match 2.8%; Score 36; DB 32; Length 162;

Best Local Similarity 42.9%; Pred. No. 1.04e-06;

Matches 54; Conservative 29; Mismatches 42; Indels 1; Gaps 1;

Db 14 tggagcgagcbgctbaggcbbgcgcbgaggtgtgtctgcgbbbtbtggagbgcgb 73

Y 9 TNGGGGCTCTCTCTCTGTGAGTCGGCGCCGCGGGCTGTGCTGTGCGACGCGC 68

Db 74 tgcgagcgcbgcbgagcgcbgtgcbtbpagbtp-cbgcbgaggtcgcgagbgpct 132

Y 69 GCGGCGACGAGCTCGGCACATGAGCGGCTTCAGCACCGAGGAGCGCGCCGCTTCT 128

Db 133 tcbtgg 138

Y 129 CCCGCG 134

RESULT 12

ID V44650 standard: DNA; 91 BP.

AC V44650;

DT 06-OCT-1998 (first entry)

DE Mammalian DNA replication origin consensus sequence, uniorconsensus;

KW DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;

KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;

KW gene therapy; ss.

PN Mammalia.

OS M09827200-A2.

PD 25-JUN-1998.

PF 12-DEC-1997; CA0972.

PR 21-MAY-1997; US-047322.

PR 16-DEC-1996; US-033374.

PA (UYMC-) UNIV MCGILL.

PI COSCONS NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;

PT WPI; 98-562770/31.

PT Human or mammalian origin of replication consensus sequences - for

PT inhibiting DNA replication, for controlling initiation of

PT replication, maintaining circular plasmids and in assembly of human

PT artificial chromosomes

PS Claim 1: Page 42: 54pp; English.

CC This sequence represents a human or mammalian DNA replication origin

CC consensus sequences of the invention, designated uniorconsensus.

CC Administration of the consensus sequence or an anti-gene (comprising a

CC double stranded copy of the consensus) is used to inhibit DNA replication

CC in vivo or in vitro. The consensus sequences can also be inserted into an

CC expression vector, used subsequently for in vitro transfection of

CC mammalian cells, to control initiation of DNA replication. They can also

CC be used to maintain circular plasmids that are capable of

CC semi-conservative replication in proliferating mammalian cells, or

CC inserted into mammalian or human artificial chromosome vectors for gene

CC therapy. Particularly, they are used to create shuttle vector constructs

CC for defining the essential mammalian elements required for maintenance of

CC chromosomal function. The consensus sequence can be combined with cloned

CC human telomeres and large centromeric blocks for assembly of human

CC artificial chromosomes and maintained as bacterial plasmids, circular or

CC linear, large or small yeast artificial chromosomes (YACS) or as episomal

CC elements.

CC Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

Query Match 2.7%; Score 34; DB 46; Length 91;

Best Local Similarity 17.1%; Pred. No. 1.29e-05;
Matches 13; Conservative 47; Mismatches 14; Indels 2; Gaps 2;
Db 8 krawrwwkda-vwagkrwwkwwhrassacmwkaaklwkqgwltwariywkgrkmmwt 66
457 TGAATTGGAAGCAAGATGTGCAAGAGTGAATATTTGCGTGAAGTCTAGAGCAT 516
QY 517 AT-TGCTATGATGTA 531
Db 67 wkawdatakwwkda 82
QY 517 AT-TGCTATGATGTA 531
RESULT 13
ID Q70469 standard; DNA; 114 BP.
AC Q70469:
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PT disclosure; page 35; 25pp; English.
PS Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC This generic formula can be represented as follows: X(TGC)(NNB)10-
CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
CC sites (X is not the same as Y) that are not specified further. This
CC sequence generates peptides that are cleavable in structure. Other
CC generic sequences are shown in Q70465-68. Other specific peptides
CC generated by these generic sequences are shown in R65150-54. TSARs are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The oligonucleotides are also designed so that the expressed peptide
CC contains 2 or 4 cysteine residues positioned in, or flanking, the
CC unpaired or variant residues. These residues confer some degree of
CC conformational rigidity to the peptides. The TSARs or compns. comprising
CC a TSAR binding domain can be used in vivo to deliver a chemically or
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC or enzyme, to the specific target or on the cell. They can also replace
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC and therefore circumvent the need for complex methods of hybridoma
CC formation or in vivo antibody production. The TSARs are easily
CC characterised and have designed activity allowing direct and rapid
CC detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
Query Match 2.7%; Score 35; DB 12; Length 114;
Best Local Similarity 4.7%; Pred. No. 3.68e-06;
Matches 5; Conservative 32; Mismatches 70; Indels 0; Gaps 0;
Db 6 bmbnbmbnbmbnbmbnbmbnbgcnbmbnbmbnbmbnbmbnbmbnnnnnnnn 65
QY 31 CGCGCGCGGCTGTGCTGTGCGAGCGCGCGCGCGAGACTCCGCGCACTAT 90

Db 66 btgcnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbt 112
QY 91 GAGCGGCTTACGACCGAGAGCGCGCGCGCGCTCTCCCTGAGCT 137
RESULT 14
ID Q70467 standard; DNA; 114 BP.
AC Q70467:
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PE 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PsDB: R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PT disclosure; page 35; 25pp; English.
PS Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpaired or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match 2.7%; Score 34; DB 12; Length 114;
Best Local Similarity 1.9%; Pred. No. 1.29e-05;
Matches 2; Conservative 33; Mismatches 71; Indels 0; Gaps 0;
Db 3 bmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbgcnbmbnnnnnnnn 62
QY 34 CGCGCGTGGCGGCTGTGCTGTGCGAGCGCGCGCGAGACTCCGCACTATGAG 93
Db 63 bmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnbmbnb 108
QY 94 CGGCTTACGACGAGAGCGCGCGCGCTCTCCCTGAGATAC 139
RESULT 15
ID Q70467 standard; DNA; 114 BP.

Page 9

Search completed: Mon Dec 27 13:07:45 1999
Job time : 830 secs.

MOLISERI
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Dec 27 13:08:02 1999; MasPar time 92.62 Seconds
1190.364 Million cell updates/sec
Tabular output not generated.

Title: >US-09-415-540-2
Description: (1-1275) from US09415540.seq
Perfect Score: 1274
N.A. Sequence: 1. CAAGAGGTGCGGCTCTCT.....GAATTAATTAATTTGCTG 1275
Comp: GTTCCACANCCCGACGAGA.....CTTATTTAATTAACGAC

Scoring table: TABLE default
Gap 6

Match STD : Dbase 0; Query 0

Searched: 165362 seqs, 43334748 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 8.783; Variance 4.614; scale 1.904

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1274	100.0	1275	3	US-08-741-	Sequence 2, Applicatio	0.00e+00
2	80	6.3	7218	2	US-08-232-	Sequence 14, Applicati	5.75e+40
3	34	2.7	965	3	US-08-388-	Sequence 22, Applicati	1.04e-07
4	28	2.2	215	1	US-08-238-	Sequence 5, Applicatio	4.80e-04
5	28	2.2	965	3	US-08-388-	Sequence 22, Applicati	1.85e-03
6	27	2.1	215	1	US-08-238-	Sequence 5, Applicatio	1.85e-03
7	24	1.9	65	1	US-08-471-	Sequence 145, Applicat	9.22e-02
8	24	1.9	68	1	US-08-471-	Sequence 143, Applicat	9.22e-02
9	24	1.9	74	4	PCT-US95-1	Sequence 100, Applicat	9.22e-02
10	24	1.9	75	4	PCT-US95-1	Sequence 99, Applicati	9.22e-02
11	24	1.9	75	4	PCT-US95-1	Sequence 99, Applicati	9.22e-02
12	24	1.9	81	4	PCT-US95-1	Sequence 98, Applicati	9.22e-02
13	24	1.9	82	4	PCT-US95-1	Sequence 97, Applicati	9.22e-02
14	24	1.9	82	4	PCT-US95-1	Sequence 97, Applicati	9.22e-02
15	23	1.8	74	4	PCT-US95-1	Sequence 94, Applicati	3.24e-01
16	23	1.8	74	4	PCT-US95-1	Sequence 94, Applicati	3.24e-01
17	23	1.8	81	4	PCT-US95-1	Sequence 92, Applicati	3.24e-01
18	23	1.8	81	4	PCT-US95-1	Sequence 98, Applicati	3.24e-01
19	23	1.8	92	3	US-08-353-	Sequence 16, Applicati	3.24e-01
20	23	1.8	396	4	PCT-US95-1	Sequence 27, Applicati	3.24e-01

21	23	1.8	396	2	US-08-334-	Sequence 27, Applicati	3.24e-01
22	23	1.8	657	1	US-08-412-	Sequence 1, Applicatio	3.24e-01
23	23	1.8	1241	1	US-07-593-	Sequence 6, Applicatio	3.24e-01
24	22	1.7	65	1	US-08-471-	Sequence 145, Applicati	1.11e+00
25	22	1.7	66	1	US-08-471-	Sequence 144, Applicati	1.11e+00
26	22	1.7	68	1	US-08-471-	Sequence 143, Applicati	1.11e+00
27	22	1.7	69	1	US-08-471-	Sequence 142, Applicati	1.11e+00
28	22	1.7	69	1	US-08-471-	Sequence 142, Applicati	1.11e+00
29	22	1.7	74	4	PCT-US95-1	Sequence 94, Applicati	1.11e+00
30	22	1.7	81	4	PCT-US95-1	Sequence 92, Applicati	1.11e+00
31	22	1.7	105	1	US-07-865-	Sequence 13, Applicati	1.11e+00
32	22	1.7	2850	3	US-08-224-	Sequence 7, Applicatio	1.11e+00
33	22	1.7	7218	2	US-08-323-	Sequence 14, Applicati	1.11e+00
34	22	1.7	9636	2	US-08-323-	Sequence 1, Applicatio	1.11e+00
35	22	1.7	26700	2	US-08-488-	Sequence 5, Applicatio	1.11e+00
36	22	1.7	26700	2	US-08-472-	Sequence 1, Applicatio	1.11e+00
37	21	1.6	59	4	PCT-US95-1	Sequence 95, Applicati	3.66e+00
38	21	1.6	66	4	PCT-US95-1	Sequence 93, Applicati	3.66e+00
39	21	1.6	1302	3	US-08-529-	Sequence 3, Applicatio	3.66e+00
40	21	1.6	1362	5	5290690-8	Patent No. 5290590.	3.66e+00
41	21	1.6	1723	3	US-08-804-	Sequence 1, Applicatio	3.66e+00
42	21	1.6	1971	5	5349058-1	Patent No. 5349058.	3.66e+00
43	21	1.6	3373	1	US-08-273-	Sequence 2, Applicatio	3.66e+00
44	21	1.6	4766	4	PCT-US93-0	Sequence 10, Applicati	3.66e+00
45	21	1.6	8174	1	US-07-914-	Sequence 5, Applicatio	3.66e+00

ALIGNMENTS

RESULT 1
ID US-08-741-437-2 STANDARD, DNA; UNC; 1275 BP.
AC xxxxxx
DE Sequence 2, Application US/08741437
CC Sequence 2, Application US/08741437
CC Patent No. 15843655
CC GENERAL INFORMATION:
CC APPLICANT: Hawkins, Phillip R.
CC APPLICANT: Hillman, Jennifer L.
CC TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSPO Version 1.5.
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/741,437
CC FILING DATE: Filed Herewith
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0148 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 855-0555
CC TELEFAX: (415) 845-4166
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1275 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA

CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE: Consensus
SO SEQUENCE 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 OTHER

Query Match	100.0%; Score 1274; DB 3; Length 1275;
-------------	--

Matches	1275;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-------	--------------	----	------------	----	--------	----	------	----

Dp	1	CAAGAGGTGNGGGCTCTCTCTCTGTAGTCGGGCGCCGCTCGGGGCTGGTGGCTCTG	60
Qy	1	CAAGAGGTGNGGGCTCTCTCTCTGTAGTCGGGCGCCGCTCGGGGCTGGTGGCTCTG	60
Dp	61	GCACGGGGGGGCGCAGAGACTCGGCACTATGAGCGGCTTAGACACGAGGAGCGCCCG	120
Qy	61	GCACGGGGGGGCGCAGAGACTCGGCACTATGAGCGGCTTAGACACGAGGAGCGCCCG	120
Dp	121	GCCCTTCTCCTTGAGATACGAGCTTCTCTAAAATGAGAAAAGCAATATATCTCC	180
Qy	121	GCCCTTCTCCTTGAGATACGAGCTTCTCTAAAATGAGAAAAGCAATATATCTCC	180
Dp	181	ATTTCATGATATTTCCATTTATGAGATTAAGAGTGTTCACATGGTATTTAGATAC	240
Qy	181	ATTTCATGATATTTCCATTTATGAGATTAAGAGTGTTCACATGGTATTTAGATAC	240
Dp	241	ACGGTGGTCTAATGCAAAAATGAGATTTGTCTACAAAGACCTTTAAACCTATTAA	300
Qy	241	ACGGTGGTCTAATGCAAAAATGAGATTTGTCTACAAAGACCTTTAAACCTATTAA	300
Dp	301	AGATGTGAAAAAGAAAACTTGGCTATGTTGGCAATTTGTTCCGTATTAAGATAT	360
Qy	301	AGATGTGAAAAAGAAAACTTGGCTATGTTGGCAATTTGTTCCGTATTAAGATAT	360
Dp	361	CTGGAACATATGTTGCCATTCCTCGACTTGAGATTGGAGAACCCGAGGACAAATAT	420
Qy	361	CTGGAACATATGTTGCCATTCCTCGACTTGAGATTGGAGAACCCGAGGACAAAT	420
Dp	421	TGGCTGTGTGGTGACAAATGCAATTTGATGTGTGTGAATTTGAAAGCAAGTATG	480
Qy	421	TGGCTGTGTGGTGACAAATGCAATTTGATGTGTGTGAATTTGAAAGCAAGTATG	480
Dp	481	AAGAGGTGAATAATTTGGCGTGAAGTTCTAGACATTTGGTATATTGACGAAGGA	540
Qy	481	AAGAGGTGAATAATTTGGCGTGAAGTTCTAGACATTTGGTATATTGACGAAGGA	540
Dp	541	AACCGATGGAAAGTCAATTCGCATTAATGAGATGATCCGATGACAGCCATTATAT	600
Qy	541	AACCGATGGAAAGTCAATTCGCATTAATGAGATGATCCGATGACAGCCATTATAT	600
Dp	601	TATCAATGATGTCAAAAGCGGTGAACCTGGCTACTTGAAGTACATGTGGAGTGT	660
Qy	601	TATCAATGATGTCAAAAGCGGTGAACCTGGCTACTTGAAGTACATGTGGAGTGT	660
Dp	661	AAGTATAGGTTCTGTATGAAAAACGAAAAATGAGTTGGCGTTTATATGACAAAT	720
Qy	661	AAGTATAGGTTCTGTATGAAAAACGAAAAATGAGTTGGCGTTTATATGACAAAT	720
Dp	721	AGATTAAGAGCTTGGCAATTTGATATTTAATAAACCATCATGACCATTTGAAACAT	780
Qy	721	AGATTAAGAGCTTGGCAATTTGATATTTAATAAACCATCATGACCATTTGAAACAT	780
Dp	781	GACTTAAGAAAAAGATGAGAAAGAAATCATGATGATATACAACTTTGTCTAGAGCC	840
Qy	781	GACTTAAGAAAAAGATGAGAAAGAAATCATGATGATATACAACTTTGTCTAGAGCC	840
Dp	841	CTTCAAGTGTGATCTGATCTCTCCAGAGCAATTTGTGATGTTTACCAACACCTGT	900
Qy	841	CTTCAAGTGTGATCTGATCTCTCCAGAGCAATTTGTGATGTTTACCAACACCTGT	900
Dp	901	ATCTGCTGCACAGTACCACAGACGTGTGATAGTGTTCATACCCAGAAAACTAAT	960
Qy	901	ATCTGCTGCACAGTACCACAGACGTGTGATAGTGTTCATACCCAGAAAACTAAT	960

Db	961	AGTTTCTCGGAAATACAACCTCAATATGCTACATCGTGTATCTGATGTATAGAG	1020
OY	961	AGATTTCCTGGAAATACAACCTCAATATGCTACATCGTGTATCTGATGTATAGAG	1020
Db	1021	TAAAGTAGTAGCTTTTCCAAGCTTTAAATTTGTAGAACCTCATCTAACGTAAAGTAAATTC	1080
OY	1021	TAAAGTAGTAGCTTTTCCAAGCTTTAAATTTGTAGAACCTCATCTAACGTAAAGTAAATTC	1080
Db	1081	TGCGTGACTAATCAATATATCAATGACAAATGTTATCCATTAAGCATTTTTCATATCTCA	1140
OY	1081	TGCGTGACTAATCAATATATCAATGACAAATGTTATCCATTAAGCATTTTTCATATCTCA	1140
Db	1141	ACTAGATTAACCTTTTAGCACATGCTTAAATATATCAAGACGTTGTCAATTTGGAACTCACTT	1200
OY	1141	ACTAGATTAACCTTTTAGCACATGCTTAAATATATCAAGACGTTGTCAATTTGGAACTCACTT	1200
Db	1201	GTGAATAGATGTGCAAGGGAGACATATTTGGATGTATATGTTACCATATATGTAGAAAT	1260
OY	1201	GTGAATAGATGTGCAAGGGAGACATATTTGGATGTATATGTTACCATATATGTAGAAAT	1260
Db	1261	AAAATTAATTTGCTG 1275	
OY	1261	AAAATTAATTTGCTG 1275	

RESULT 2
ID US-08-32-463-14 STANDARD; DNA; UNC; 7218 BP
AC xxxxxx
DT
DE Sequence 14, Application US/08222463
CC Patent 14, Application US/0832463
CC Patent No. 5670367

```

CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232,463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935,313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear

```

CC IMMEDIATE SOURCE:
CC CLONE: PTZgpt-Fls
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 6.3%; Score 80; DB 2; Length 7218;
Best Local Similarity 0.8%; Pred. No. 5,75e-40;
Matches 3; Conservative 217; Mismatches 140; Indels 0; Gaps 0;

Db 1064 GATYY 1123
|||.....
Cc 808 GATTCCTTTTCATTCGTTTCTTACTACATAGCTTCCATGAGTGGCTTTT 749
Db 1124 YY 1183
Cc 748 AATAATATCAATGCGCAATCTCTATCTTAATTCGTGATTAACGCAACTATTTTC 689
Db 1184 YY 1243
Cc 688 TGGTTTTCATCAGAACCTTATACCTTCTAACAACGACGAGTGGCTTAAAGTAGCC 629
Db 1244 YY 1303
Cc 628 AGCTTACACCGCTTGACATCATATGATATCATATATGCTGATCAGATCATCCAC 569
Db 1304 YY 1363
Cc 568 ATTAATGCAATGACTTTCAGCTGCTTCCCTTGTGTAATCATAGCCAAATAGCTTAG 509
Db 1364 YY 1423
Cc 508 AACTTTCACGCCAATATATTCACCTTCTCACAATACCTTGTCCAAATTCACACACATC 449

RESULT 3
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC xxxxxx

DE Sequence 22, Application US/08388672A
CC Sequence 22, Application US/08388672A
CC Patent No. 5795961

GENERAL INFORMATION:
CC APPLICANT: Wallace, T. Paul
CC APPLICANT: Harris, William J.
CC APPLICANT: Carr, Frank J.
CC APPLICANT: Old, Lloyd J.
CC APPLICANT: Weilt, Sydney
CC APPLICANT: Kitamura, Kunio
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B
CC TITLE OF INVENTION: Antibodies
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Felle and Lynch
CC STREET: 805 Third Avenue
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10022

COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/388,672A
CC FILING DATE: 14-FEB-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, No. 5795961man D.
CC REGISTRATION NUMBER: 30,946
CC REFERENCE/DOCKET NUMBER: LUD 5409
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-688-9200
CC TELEFAX: 212-838-3884

CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 2.7%; Score 34; DB 3; Length 965;
Best Local Similarity 12.9%; Pred. No. 1.04e-07;
Matches 15; Conservative 62; Mismatches 37; Indels 2; Gaps 2;

Db 826 YNGKRRVYMATSSRSRSYTAADTAVYCYRGSRYSDDGDYMGCT-TYVYSSHYVD 884
|||.....
Cc 8 TTTGGGGCTCTCTCTGTCAGTCGGCGCGGCTGGGCTGTGCTGTGGCAGCGG 67
Db 885 MTSSASAVGDRVTCRSSTHNGNTYMYGKAKRYRNRSNGVSRSGSGGSDTY 940
Cc 68 CGCGCG-CAGGACTCGGCACATATAGCGGCTTCACGACGAGAGCGCGCGCGC 122

RESULT 4
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC xxxxxx

DE Sequence 5, Application US/08238163
CC Sequence 5, Application US/08238163
CC Patent No. 5569830

GENERAL INFORMATION:
CC APPLICANT: BENNETT, Alan
CC APPLICANT: LABAVITCH, John M.
CC APPLICANT: POWELL, Ann
CC APPLICANT: STOLTZ, Henrik
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL
CC TITLE OF INVENTION: POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEA
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend Kourile and Crew
CC STREET: Stewart Street Tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493

COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/238,163
CC FILING DATE: 03-MAY-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc_feature
CC LOCATION: 1..215
CC OTHER INFORMATION: /standard name= "Deduced amino acid
CC OTHER INFORMATION: sequence of PCIP from bean."
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 2.2% Score 28; DB 1; Length 215;
Best Local Similarity 13.3%; Pred. No. 4,806-04;
Matches 20; Conservative 62; Mismatches 67; Indels 1; Gaps 1;

Db 1 MTNVTMSSTVSRASNDKAKKDGNT-TSSWTTDCNRWGVCDPDTTRVNNDSHN 59
22 CTGTGATCGCGCCGCGGCTGTGTGCTGTGTGCGAGCGCGCGCGAGACTC 81
Db 60 KYSSANVYCGNNVCAKTHYTHNTVNSGADSKVTDSYNSNGTGDNRSGADS 119
82 CGGACATATAGCGCGCTCAGACCGAGAGCGCGCGCGCTTCTCCGAGATACG 141
Db 120 YGSKRTAMTSRNRGTAKNNAVDSRNGDA 149
142 AGCTCTCTCAAAATGAGAAAGCAATA 171

RESULT 5
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC xxxxxx

DE Sequence 22, Application US/08388672A
CC Sequence 22, Application US/08388672A
CC Patent No. 5795961
CC GENERAL INFORMATION:
CC APPLICANT: Wallace, T. Paul
CC APPLICANT: Harris, William J.
CC APPLICANT: Carr, Frank J.
CC APPLICANT: Old, Lloyd J.
CC APPLICANT: Well, Sydney
CC APPLICANT: Kitamura, Kunio
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Felfe and Lynch
CC STREET: 805 Third Avenue
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10022
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/388,672A
CC FILING DATE: 14-FEB-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, No. 5795961man D.
CC REGISTRATION NUMBER: 30,946
CC REFERENCE/DOCKET NUMBER: LUD 5409
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-688-9200
CC TELEFAX: 212-838-3884
CC INFORMATION FOR SEQ. ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 2.2% Score 28; DB 3; Length 965;
Best Local Similarity 13.6%; Pred. No. 4,806-04;
Matches 25; Conservative 83; Mismatches 74; Indels 2; Gaps 2;

Db 780 KCUHVHVHSGVRSRSTASDYTTSYWGVYRGWMDYGGGYNYNCKRGRTVMDATS 839

Cp 1003 TGAACAGATGTAGCATATACACTTGTATTC-AGAGAAATCATAGTTTCTGCT 945

Db 840 SNNRSVTAADTAIVYCVGRSVDSDGDYWGTTVYSSHVKDMTSSSASVGRVTT 899
944 GATGAGAACACTATATCACGCTGTGTACTGTGACGAGCATTCACAGGGGTGGTGA 885

Db 900 CRSTTGNNTYTWYKRAKRYVSNSSVSR-SGSSSGDYTTSSDPAIVYCGTHARTG 958
884 AAGCATCACAAATGCTCTGCGACATCAGATCATTGAAGGGGCTCTCAGACAAG 825

Db 959 TKYK 962
Cp 824 TTGT 821

RESULT 6
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC xxxxxx

DE Sequence 5, Application US/08238163
CC Sequence 5, Application US/08238163
CC Patent No. 5569830
CC GENERAL INFORMATION:
CC APPLICANT: BENNETT, Alan
CC APPLICANT: LABAVITCH, John M.
CC APPLICANT: POWELL, Ann
CC APPLICANT: STORZ, Henrik
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Townsend and Townsend Kourie and Crew
CC STREET: Stuart Street Tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/238,163
CC FILING DATE: 03-MAY-1994
CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ. ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 1..215
CC OTHER INFORMATION: /standard_name="Deduced amino acid
CC OTHER INFORMATION: sequence of p53p from bean."
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 2.1% Score 27; DB 1; Length 215;
Best Local Similarity 15.0%; Pred. No. 1,856-03;
Matches 25; Conservative 65; Mismatches 75; Indels 2; Gaps 2;

Db 39 RTWGVCDTDTT-YRVNDSGHNKYSSANYNYGNNVGAKTHYTHNTVNSGADSKTYTDS 97

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mistrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 100:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 74 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 74 BP: 6 A; 6 C; 1 G; 1 T; 60 OTHER.
SO

Query Match 1.9%; Score 24; DB 4; Length 74;
Best Local Similarity 8.7%; Pred. No. 9.22e-02;
Matches 6; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

Db 6 VNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN 65
Cc 74 GCCCGCCGCCGCGCCACAGACGACGACCCCGCGCGCGACTGACACAGAGAGAG 15
Db 66 ACCACACACC 74
Cc 14 CCCCNAACC 6
Cp

RESULT 10
ID PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC xxxxxx
DT

Sequence 99, Application PC/TUS9511934
GENERAL INFORMATION:
CC APPLICANT: Cytogen Corporation
CC TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC NUMBER OF SEQUENCES: 103
CC CURRENT APPLICATION DATA:
CC CORRESPONDENCE ADDRESS:
CC ADDRESSER: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/11934
CC FILING DATE: 20-SEP-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mistrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 99:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 75 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 75 BP: 1 A; 1 C; 7 G; 5 T; 61 OTHER.
SO

Query Match 1.9%; Score 24; DB 4; Length 75;
Best Local Similarity 7.7%; Pred. No. 9.22e-02;
Matches 5; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

Db 4 AGNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNN 63
Cc 3 AGAGCTTNGGGGCTCTCTGTCATGCGCGCGCGGCTGTGGCTGTGTGGC 62
Db 64 NNBGG 68
Cc 63 AGCGG 67
Cp

RESULT 11
ID PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC xxxxxx
DT

Sequence 99, Application PC/TUS9511934
GENERAL INFORMATION:
CC APPLICANT: Cytogen Corporation
CC TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC NUMBER OF SEQUENCES: 103
CC CURRENT APPLICATION DATA:
CC CORRESPONDENCE ADDRESS:
CC ADDRESSER: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/11934
CC FILING DATE: 20-SEP-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mistrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 1101-196-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 99:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 75 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 75 BP: 1 A; 1 C; 7 G; 5 T; 61 OTHER.
SO

Query Match 1.9%; Score 24; DB 4; Length 75;
Best Local Similarity 7.5%; Pred. No. 9.22e-02;
Matches 5; Conservative 20; Mismatches 42; Indels 0; Gaps 0;

Db 5 GNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNN 64
Cc 442 GTCATTGTCACCAACAGCAGATGTTATGTCAGCTGGCTTCCCAAGTCTG 383
Db 65 NNBGTTG 71
Cc 382 AGGATG 376
Cp

RESULT 12
ID PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.
AC xxxxxx
DT

```

AC      xxxxxx
DT      Sequence 98, Application PC/TUS9511934
DE      Sequence 98, Application PC/TUS9511934
CC      GENERAL INFORMATION:
CC      APPLICANT: Cytogen Corporation
CC      TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC      TITLE OF INVENTION: Peptide Libraries
CC      NUMBER OF SEQUENCES: 103
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Pennie & Edmonds
CC      STREET: 1155 Avenue of the Americas
CC      CITY: New York
CC      STATE: New York
CC      COUNTRY: USA
CC      ZIP: 10036
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US95/11934
CC      FILING DATE: 20-SEP-1995
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Mistrock, S. Leslie
CC      REGISTRATION NUMBER: 18,872
CC      REFERENCE/DOCKET NUMBER: 1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (212) 790-9090
CC      TELEFAX: (212) 869-9741/8864
CC      TELEX: 66141 PENNIE
CC      INFORMATION FOR SEQ ID NO: 98:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 81 base pairs
CC      TYPE: nucleic acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: DNA (genomic)
CC      SEQUENCE 81 BP; 6 A; 6 C; 4 G; 5 T; 60 OTHER.
SQ
Query Match      1.9%; Score 24; DB 4; Length 81;
Best Local Similarity 8.7%; Pred. No. 9.22e-02;
Matches      6; Conservative      20; Mismatches 43; Indels 0; Gaps 0;

Db      13 VNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN
Cp      74 GCCCGCCCGCGCTGCCACAGACGACACCCGCCGACGGCGCGACCTGCACAGAGAGAG 15
Db      73 ACCACACACC 81
          || ||||
Cp      14 CCCCNAAACC 6

RESULT      13
ID      PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC      xxxxxx
DT      Sequence 97, Application PC/TUS9511934
DE      Sequence 97, Application PC/TUS9511934
CC      GENERAL INFORMATION:
CC      APPLICANT: Cytogen Corporation
CC      TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC      TITLE OF INVENTION: Peptide Libraries
CC      NUMBER OF SEQUENCES: 103
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Pennie & Edmonds
CC      STREET: 1155 Avenue of the Americas
CC      CITY: New York
CC      STATE: New York
CC      COUNTRY: USA
CC      ZIP: 10036

```

[illegible]

US-09-415-540-2.rni

Page 8

QY 3. AGAGGTTTNGGGCTCTCTTCCTTGTCAGTCGGCGCGCGGTGGGGCTAGTGGCTCTGTGGC. 62
Db 64 NNENACG. 70
QY 63. AGCGGCG. 69

Search completed: Mon Dec 27 13:12:12 1999
Job time : 250 secs.

Age Group	U.S. should take action (%)	U.S. should not take action (%)
18-29	85	15
30-49	82	18
50-69	88	12
70+	92	8